SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(139TH Edition)

JULY 2024

BY: CENTRAL LIBRARY

DELHI TECHNOLOGICAL UNIVERSITY
(FORMERLY *DELHI COLLEGE OF ENGINEERING)*
GOVT. OF N.C.T. OF DELHI
SHAHBAD DAULATPUR, MAIN BAWANA ROAD
DELHI 110042

PREFACE

This is the **One Hundred Thirty Ninth** Issue of Current Awareness Bulletin started by Delhi Technological University, Central Library. The aim of the bulletin is to compile, preserve and disseminate information published by the faculty, students and alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of Delhi Technological University (DTU) as a whole to the academia.

The bulletin contains information resources available in the internet in the form of articles, reports, presentations published in international journals, websites, etc. by the faculty and students of DTU. The publications of faculty and student which are not covered in this bulletin may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose.

The learned faculty and students are requested to provide their uncovered publications to the library either through email or in CD, etc. to make the bulletin more comprehensive.

This issue contains the information published during **July, 2024**. The arrangement of the contents is alphabetical. The full text of the article which is either subscribed by the university or available in the web is provided in this bulletin.

Central Library

CONTENTS

1. A Comparative Study of Machine Learning Models for Network Traffic Classification in the IoT Landscape, ***8.Divyanshu Thakur***, ***8.Srijan Sah***, ***8.Priyansh Ailsinghani*** and ***3.Virender Ranga***, IT, DTU

2. A fexible memory device made of SnO2-hBN nanocomposite exhibits stable resistive switching application, ***6.Km. Komal***, ***3.Mukhtiyar Singh*** and ***3.Bharti Singh***, Applied Physics, DTU

3. A NEW DIFFERENTIAL SUBORDINATION TECHNIQUE FOR A SUBCLASS OF STARLIKE FUNCTIONS, ***3.S. SIVAPRASAD KUMAR*** AND ***6.POOJA YADAV***, Applied Mathematics, DTU

4. A novel hybrid intelligent approach for solar photovoltaic power prediction considering UV index and cloud cover, ***6.Rahma Aman***, ***3.M. Rizwan*** and Astitva Kumar, Electrical, DTU

5. A novel investigation on single-loop pulsating heat pipe filling with hybrid nanofluids: numerical and experimental, ***6.Prem shanker yadav***, Jitendra sharma, Mohd Hussain, Imran ahmed Khan, Kartik Goyal and Samer Fikry Ahmed, Mechanical, DTU

6. A uniformly convergent method for two-parameter singularly perturbed parabolic partial differential equations with a large shift, ***3.Aditya Kaushik***, ***6.Nitika Sharma***, Manju Sharma, Monika Arora and ***6.Shivani Jain***, Applied Mathematics, DTU

7. Addressing Health Disparities through Community-Based Public Health Initiatives, ***6.Naima Tiwari***, Public Health, DTU

8. An experimental study of game theory with various word embeddings for automatic extractive text summarization, ***3.Minni Jain***, ***3.Rajni Jindal*** and Amita Jain, CSE, DTU

9. An Implementation of a Frequency Decomposition Algorithm for Power Quality Smoothening in a Hybrid SPVA-BES Microgrid, ***7.1, Divyank Srivastava***, Vivek Narayanan, Bhim Singh, and Arunima Verma, Power Systems, DTU

10. Analysis of PCE-based Superplasticiser for the Different Types of Cement using Marsh Cone Test, Rahul Kumar, Manvendra Verma and ***3.Nirendra Dev***, Civil, DTU

11. Analyzing Yoga Pose Recognition: A Comparison of MediaPipe and YOLO Keypoint Detection with Ensemble Techniques, ***7.Debashree Debalaxmi***, ***3.Dinesh Kumar Vishwakarma*** and ***3.Virender Ranga***, IT, DTU

12. Assessing Gender Bias in Machine Translation, ***7.Rishabh Jain***, CSE, DTU

13. Barrow holographic dark energy model in f (R,T) theory, Kanchan Devi, ***6.Ajay Kumar*** and Pankaj Kumar, Applied Mathematics, DTU

14. Base and Exponent Prediction in Mathematical Expressions using Multi-Output CNN, ***8.Md Laraib Salam***, Akash S Balsaraf and Gaurav Gupta, Electrical, DTU

15. Camera-based Mobility Framework for Visually Impaired Pedestrians in Unstructured Environments, Aditya Kumar, Anwesha Chakravarty, ***6.Ayesha Choudhary*** and ***3.S. Indu***, ECE, DTU

16. Carboxymethyl Guar Gum-Based Bioactive and Biodegradable Film for Food Packaging, ***6.Tanvi Singh***, ***6.Meenakshi Tanwar*** and ***3.Rajinder K. Gupta***, Chemistry, DTU

17. Classification of Banana Plant Leaves Based on Nutrient Deficiency Using Vision Transformer, ***6.Sumit Kumar Sharma*** & ***3.Prof. Dinesh Kumar Vishwakarma***, IT, DTU

18. Comparative Analysis of various Slot in Patch Antennas: A Review, ***3.Sumit Kumar Khandelwal***, ***8.1Anupam Verma, 8.1Amarnath Yadav*** and ***8.1Arpan Goyal, ECE, DTU***

19. Complementary Dynamics of Global and Local Centrality for Influence Maximization in Social Networks, ***3.Rajesh Kumar Yadav***, ***8.Aaryan Arora***, ***8.Aayush Patel*** and ***8.Aditya Gupta***, CSE, DTU

20. Cosmological tests of the dark energy models in Finsler-Randers Space-time, Z. Nekouee, ***7.Himanshu Chaudhary***, S. K. Narasimhamurthy, S. K. J. Pacif and Manjunath Malligawad, Applied Mathematics, DTU

21. Curcumin and ZnO-NPs loaded xanthan gum and hydroxypropyl guar gum-based hydrogel film for food packaging, ***8.Ishita Parmar***, ***6.Meenakshi Tanwar*** and ***3.Rajinder K. Gupta***, Applied Chemistry, DTU

22. Design and theoretical study of rectangular photonic crystal fiber based sensor for chemical sensing in terahertz regime, ***8.1 D Garg, 8.1 J Singh,*** and ***3.A Kumar,*** Applied Physics, DTU

23. Document Information Retrieval with Deep Learning, ***3.Abhilasha Sharma*** and ***7.Aparna Arya***, Software Engineering, DTU

24. DoSLex: automatic generation of all domain semantically rich sentiment lexicon **3.** ***Minni Jain, 3. Rajni Jindal*** ,Amita Jain,CSE, DTU

25. Ecological Representation in the Maithili Songs of Vidyapati, ***6.Mallika Tosha*** and ***3.Dr. Rajiv Ranjan Dwivedi***, Humanities, DTU

26. Effect of various energy storage phase change materials (PCMs) and nano-enhanced PCMs on the performance of solar stills: A review ***6.Aseem Dubey*** and ***3.Akhilesh Arora***, Mechanical, DTU

27. Efficient hashing technique for malicious profle detection at hypervisor environment, Anumukonda Naga ***6.Seshu Kumar***, ***3.Rajesh Kumar Yadav*** and ***3.Nallanthighal Srinivasa Raghava***, CSE and Electronics, DTU

28. Enhancement in sugar extraction from Pistia stratiotes through statistical optimization of alkaline pre-treatment and enzymatic hydrolysis, ***6.Sweeti Mann***, ***3.Jai Gopal Sharma*** and Rashmi Kataria, Biotechnology, DTU

29. Enhancing concrete durability and strength: An innovative approach integrating abrasion and cement slurry treatment for recycled coarse aggregates, ***6.Harish Panghal*** and ***3.Awadhesh Kumar***, Civil, DTU

30. Evaluation of the role of process parameters of CMT welding and their statistical behaviour on the tensile properties of weld joint of novel aluminium hybrid composite, ***6.Vibhu Singh***, ***3.M. S. Niranjan*** and ***3.Qasim Murtaza***, Mechanical, DTU

31. Evaluation of Tribological and Cooling Performance of TiN and DLCCoated Pistons for Miniature Stirling Cryocooler ***8.M.Singhal***, R. Kumar, R.S.Walia,,S. K. Pandey, Mechanical Engineering,DTU

32. Experimental analysis of GMAW welded AA7475 T7351 alloy joints, Hari Shanker, ***3.Reeta Wattal***, Mechanical Engineering ,DTU

33. "Explainable AI" Disease Detection with Reasoning, ***3.Manoj Kumar***, ***8.Kanishk Ramrakhiyani*** and ***8.Harsh Garg***, CSE, DTU

34. Exploring the potential of dual-metallized PET towards improving the efficiency of outermost reflective layer in fire proximity clothing, Shivangi Dwivedi, ***3.Richa Srivastava*** and Prasun Kumar Roy, Applied Chemistry, DTU

35. Exploring the wear behaviour of Al6351-Gr composite reinforced with B4C, TiC, and WC particles, Shahazad Ali, ***3.Qasim Murtaza***, Pallav Gupta, Mechanical, Production and Industrial Engineering Department, DTU

36. Geographical variability of ultrafine particle concentrations in urban and background regions in India, ***6.Vignesh Mohan***,Vijay Kumar Soni, ***3.Rajeev Kumar Mishra,*** Environmental Engineering, DTU

37. Ge-Se-Te based Penrose Photonic Quasi-Crystal fiber for SCG covering 2-21 μm MIR Regime, ***6.Akash Khamaru*** and ***3.Ajeet Kumar***, Applied Physics, DTU

38. Hate Speech Detection using CoT and Post-hoc Explanation through Instruction-based Fine Tuning in Large Language Models, ***8.Palak Gupta***, ***8.Nikhilesh Jain*** and ***3.Aruna Bhat***, CSE, DTU

39. Impact of flexible vegetation on hydraulic characteristics in compound channels with converging floodplains, ***6.Noopur Awasthi, 3.Munendra Kumar*** & ***3.Ritu Raj***, Civil, DTU

40. Impact of Health Insurance on Accessibility, Utilization of Inpatient Care, and Financial Risk Protection Across States/Union Territories of India, ***6.Aashima***, ***6.Mehak Nanda*** and Rajesh Sharma, USME, DTU

41. Improved optical characteristics in BaBi2Nb2O9 ferroelectric ceramic infused with transition metal ion (W6+) and rare earth ions (Er3+/Yb3+), ***6.Ankita Banwal***, ***6.Bhavya Kumar***, Manoj Verma, Ankur Shandilya, ***3.Bharti Singh*** and ***3.Renuka Bokolia***, Applied Physics, DTU

42. Improving Intrusion Detection in a Software Defined Network using Hybrid CNN and Bi-LSTM, ***8.Ashutosh Bhattarai***, ***8.Unmik Gyawali***, ***6.Abhishek Verma*** and ***3.Virender Ranga***, IT, DTU

43. Innovative Channel Estimation Method for Rayleigh, Rician, and Nakagami Channels: The Modified M-Estimator-I, Prakhar Hari, Sanjay Mathur, Prachi Mishra, ***6.Pooja Singh*** and Hari Kumar Singh, Electrical, DTU

44. Integration of Deep Learning and Plant Biosecurity Toward Sustainable Agriculture: A SWOT Analysis, ***3.Abhilasha Sharma*** and ***6.Parul Sharma***, Software, DTU

45. Investigating the effect of scaling and temperature on the performance of improved junctionless nanowire FET through simulation analysis***, 3.Bharti,*** Applied Physics,DTU ***and 3.Poornima Mittal,*** ECE, DTU

46. Iris and Palmprint Multimodal Biometric Recognition using Novel Preactivated Inverted ResNet and Hybrid Metaheuristic Optimized DenseNet, ***3.Indu Singh***, ***8.Shruti Khatri***, ***8.Sunaina Luthra***, ***8.Tanvi Singh*** & Gunbir Singh Baveja, CSE, DTU

47. Mechanical characterization of Al6351 reinforced with TiC, B4C, and WC: Proportional effects and microstructural analysis, Shahazad Ali,***3.Qasim Murtaza***, and Pallav Gupta, Mechanical Engineering,DTU

48. Motor Imagery Features Selection with Grey Wolf Optimization for Brain Computer Interface, Pawan, Rohtash Dhiman, Jasbir Singh Saini, Ankur Saxena and ***6.Vikas Sharma***, IT, DTU

49. Multifaceted Discrete Emotion Recognition from EEG Physiological Signals via Machine Learning Techniques, ***8.Amriti Gupta***, Software Engineering, ***8.Kunal Nayyar***, ECE, ***8.Harshit Chopra*** Computer Engineering, ***6.Neha Gahlan***, Software Engineering, ***3.Divyashikha Sethia***, Software Engineering, DTU

50. Municipal organic solid waste management in the concept of urban mining and circular economy: a model from Palestine, ***6.Majed Ibrahim Al-Sari*** and ***3.A. K. Haritash***, Environmental, DTU

51. Observational Constraints on the Parameters of Hoˇrava–Lifshitz Gravity, ***6.Himanshu Chaudhary***, , Ujjal Debnath, Shibesh Kumar Jas Pacif, Niyaz Uddin Molla, Ghulam Mustafa, and Sunil Kumar Maurya, Applied Mathmatics, DTU

52. Performance analysis of fractional-order modified SRF PLL under grid abnormalities, ***6.Oinam Lotika Devi*** and ***3.Alka Singh***, Electrical, DTU

53. Phytoremediation of Chromium (VI)-Contaminated Soil by Euphorbia tithymaloides L. and Metagenomic Analysis of Rhizospheric Bacterial Community, ***6.Deepika*** and ***3.Anil Kumar Haritash***, Environmental, DTU

54. Plasma-assisted carbon nanotube for solar cell application, ***7.Suraj Kumar Singh***, ***7.Ishu Sharma*** and ***3.Suresh C. Sharma***, Applied Physics, DTU

55. Predicting Cardiovascular Disease with Machine Learning: A Comparative Analysis of Classification Models, ***7.Diya Malhotra*** and ***7.Aarushi Saini***, Applied Mathematics, DTU

56. Predictive design of KSnI3-based perovskite solar cells using SCAPS and machine learning model, ***6.Rahul Kundara*** and ***3.Sarita Baghel***, Applied Physics, DTU

57. Primary Microplastics in the Ecosystem: Ecological Effects, Risks, and Comprehensive Perspectives on Toxicology and Detection Methods, ***6.Pooja Singh***, ***6.Gunjan Varshney*** and ***3.Raminder Kaur***, Applied Chemistry, DTU

58. Quadrupole mode plasmon resonance enabled dual-band metamaterial for refractive index sensing application, ***6.Ankit***, Monu Nath Baitha, ***3.Kamal Kishor*** and ***3.Ravindra Kumar Sinha***, Applied Physics, DTU

59. Quantum Computational and Spectroscopic Investigation of 3-aminosalicylic Acid, Shahid Husain, Nupur Pandey, Nisha Fatma, Sanjay Pant and ***3.Mohan Singh Mehata***, Applied Physics, DTU

60. Quantum Resistant Blockchain-based Architecture for Secure Medical Data Sharing, ***8.Kanupriya Jain***, ***8.Mahima Singh***, ***8.Harshita Gupta*** and ***3.Aruna Bhat***, CSE, DTU

61. Quitting silently: A longitudinal research on the impact of workplace conflict and nonviolent work behavior, ***6.Suhans Bansal*** and ***3.Naval Garg***, USME, DTU

62. Real-time investigation of grid-interactive EV charger with two-stage bidirectional converter under wide voltage range scenarios, ***8.Gaurav Yadav***, ECE, ***3.Mukhtiar Singh***, Electrical Engineering, DTU

63. Rheological Behaviour Analysis of Prepared NanoCutting Fluids from Al2O3 and Tio2, Anurag Sharma, ***3.R. C. Singh***, ***3.Ranganath M. Singari*** and S. L. Bhandarkar, Mechanical and Design, DTU

64. Second generation current conveyor based capacitorless floating memristor emulator, ***6.Navnit Kumar, 3.Manjeet Kumar*** ECE, ***3.Neeta Pandey*** ECE, DTU

65. Sentiment Analysis of Unified Payments Interface (UPI) Payment App Review using Machine Learning Algorithm, ***8.Kunal Jaluthria***, ***8.Karandeep Singh***, ***8.Manik Sharma***, ***3.Rajni Jindal***, Computer Science and Engineering, DTU

66. Shapley Explainable Deep Learning based Knowledge Distillation Framework for Student's Performance Prediction, ***3.Rahul*** and ***3.Rahul Katarya***, CSE, DTU

67. Size Optimization of Grid-Tied Hybrid Energy System by Employing Forecasted Meteorological Data, P. Anand, ***6.B. Sharma*** and ***3.M. Rizwan***, Electrical, DTU

68. Strategic Pseudo-Goal Perturbation for Deadlock-Free Multi-Agent Navigation in Social Mini-Games, ***8.Abhishek Jha***, ***8.Tanishq Gupta***, ***8.Sumit Singh Rawat*** and ***3.Girish Kumar***, Mechanical, DTU

69. The collisional study of EIC waves in a magnetized dusty plasma with the inclusion of a DC electric field, ***6.Anshu***, Jyotsna Sharma and ***3.Suresh C Sharma***, Applied Physics, DTU

70. TILING SPACES AND THE EXPANDING UNIVERSE: BRIDGING QUANTUM MECHANICS AND COSMOLOGY, ***8.Aman Yadav*** and Marcelo Maciel Amaral, Civil, DTU

71. Towards optimal score level fusion for adaptive multi-biometric authentication system, ***6.Kavita***, ***3.Rajesh Rohilla*** and Gurjit Singh Walia, Electronics, DTU

72. UDCT: lung Cancer detection and classifcation using U-net and DARTS for medical CT images, Aakanksha Gupta, Ashwni Kumar and ***6.Kamakshi Rautela***, ECE, DTU

73. Utilising NLP for Enhanced Clinical Text Mining, ***8.Gaurav Gulati*** and ***8.Gaurav Upreti***, CSE, DTU

74. VEDANet: A dense blocked network for visual emotion analysis in multimedia retrieval, Kirti Sharma, Rainu Nandal, ***3.Shailender Kumar*** and Kamaldeep Joshi1, CSE, DTU

75. VRFNet-ASLiT: Fused Deep CNN and Adaptive Super Resolution Transform Based Hand Gesture Recognition, ***6.Roli Kushwaha***, ***3.Manjeet Kumar*** and ***3.Dinesh Kumar***, Electronics, DTU

76. Workplace gratitude studies: Review and future agenda using bibliometric analysis, ***3.Naval Garg*** and ***6.B. K. Punia***, USME, DTU

1. *Vice Chancellor*
2. *Pro Vice Chancellor*
3. *Faculty*
4. *Teaching-cum-Research Fellow*
5. *Asst. Librarian*
6. *Research Scholar*
7. *PG Scholar*
8. *Undergraduate Student*

1.1. *Ex Vice chancellor*
2.1. *Ex Pro Vice Chancellor*
3.1. *Ex Faculty*
4.1. *Alumni*
5.1 *Others*
6.1. *Ex Research Scholar*
7.1. *Ex PG Scholar*
8.1. *Ex Undergraduate Student*

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575268

A Comparative Study of Machine Learning Models for Network Traffic Classification in the IoT Landscape

Divyanshu Thakur
Department of Information Technology
Delhi Technological University New Delhi, India
divyanshuthakur91@gmail.com

Srijan Sah
Department of Information Technology
Delhi Technological University New Delhi, India
srijansah6@gmail.com

Priyansh Ailsinghani
Department of Information Technology
Delhi Technological University New Delhi, India
priyanshailsinghan@gmail.com

Virender Ranga
Department of Information Technology
Delhi Technological University New Delhi, India
virenderranga@dtu.ac.in

***Abstract*— Network security is a top priority in today's digital economy, where intrusions grow in sophistication and size. To safeguard network infrastructure and data, it is critical to discern between legitimate and malicious network traffic. Machine learning classification models have proven useful in this task, with decision trees, random forests, and support vector machines among the most common options. This study analyzes six machine learning techniques for determining whether network data is benign or malicious. The performance evaluation is based on the accuracy, precision, F1-score, AUC ROC score, and computational efficiency of the latest IoT dataset, the TII-SSRC-23 Dataset from Kaggle. The findings shed light on each technique's strengths and drawbacks, as well as recommendations for choosing the best approach for network traffic classification.**

Keywords*— *Machine Learning, Benign Traffic, Malicious Traffic, Classifiers, Attacks, Dataset

I. INTRODUCTION

In today's hyperconnected environment, IoT gadget proliferation is a widespread trend. However, increased connection leaves organizations and individuals vulnerable to cyber dangers. Despite the availability of tools and strategies for improving security and detecting aberrant online behavior, attackers continue to violate confidentiality and compromise sensitive data. The rapid proliferation of varied IoT devices and applications complicates the network, resulting in varying capabilities, communication protocols, and data formats.[1] [2]. Machine learning-based intrusion detection systems have gained importance for adapting to this evolving IoT landscape.

Machine learning algorithms analyze massive network data and traffic patterns to automate real-time threat detection and adaptation to evolving threats. This research focuses on developing a robust solution for network traffic classification using machine learning models to improve accuracy and adaptability in network security. Malware assaults occur when dangerous software infiltrates computer systems, networks, and devices, appearing as viruses, worms, trojans, ransomware, spyware, and adware. In cybersecurity, data is divided into benign and harmful categories. Audio, background, text, and video are examples of benign data that is both safe and legal. They are essential for authorized activities. In contrast, malicious data with detrimental intent includes brute force assaults (DNS, FTP, SSH), Denial of Service attacks (ACK, FIN, HTTP), information collecting, and Mirai (Malware).[2] Cybercriminals use malicious data to compromise system integrity, confidentiality, and availability, often aiming to gain unauthorized access, steal data, disrupt services, or extort victims.

Even with multiple safeguards, malware threats can breach a user's system, requiring machine learning-based intrusion detection systems using data-driven modeling and pattern recognition for defense.[3] Relevant machine learning classifiers for network traffic classification include:

A. Support Vector Machines (SVM):

SVMs are employed in malware analysis for binary classification tasks, with the goal of establishing the appropriate boundary between malicious and benign samples. They excel at separating complex data and can categorize malware samples based on a variety of criteria, such as opcode sequences or API call patterns. SVM is suitable for binary classification because it seeks to find the optimal hyperplane that maximizes the margin between two classes, effectively enhancing class separability[4] [5].

B. Random Forest

Random Forest is frequently used for feature selection and classification in malware investigation. It can distinguish between benign and malicious files using a variety of features including as file behavior, code characteristics, and API calls.[6] Its ensemble nature helps reduce false positives and negatives, making it suited for detecting new and emerging malware strains. Random Forest is suitable for binary classification because it combines multiple decision trees to enhance accuracy and reduce overfitting.[6]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:36:59 UTC from IEEE Xplore. Restrictions apply.

C. XGBoost

XGBoost's robust performance and ability to handle imbalanced datasets make it valuable for malware classification tasks. It can process various types of data, such as static file features and dynamic behavioural analysis results, to detect malware with high accuracy. XGBoost is suitable for binary classification because it utilizes gradient boosting to optimize a loss function, enhancing prediction accuracy and handling diverse data effectively. [7]

D. Multilayer Perceptron (MLP)

MLPs, a form of neural network, are used for deep learning in malware research. It can discover sophisticated patterns and relationships from vast datasets[8], [9]. MLP models are used for tasks such as malware family classification, which involves recognizing similarities and differences between malware samples. MLP is suitable for binary classification because it uses multiple layers of neurons to learn complex patterns and non-linear decision boundaries from data.[10]

E. Decision Tree

Decision trees are useful in malware research because they help choose traits and explain them. They can help to visualize the decision-making process and discover key markers of hostile behavior. Decision trees improve the overall performance of models using ensemble techniques like Random Forest. A decision tree is suitable for binary classification because it recursively splits data into homogenous subsets, making it intuitive and effective for distinguishing between two classes.[11]

F. Extra Trees

Extra trees are useful for feature selection and classification tasks. It operates by generating several decision trees using randomized feature selection, making it resistant to overfitting. Extra Trees is commonly used to detect and categorize malware based on various static and dynamic data, such as file properties, code behavior, and API call patterns. Extra Trees is suitable for binary classification because it builds multiple de-correlated trees using random splits and samples, which improves accuracy and robustness against overfitting. [12]

Selecting the appropriate technique involves considering various critical issues. For instance, the research objective is to categorize network traffic as benign or malicious, which requires techniques capable of handling binary classification tasks. The nature of the data, such as its volume, complexity, and mix of categorical and numerical features, necessitates techniques robust to such characteristics. Scalability and efficiency are crucial due to potentially large datasets, as indicated by the hardware specifications mentioned in the paper. Model interpretability is essential for understanding cybersecurity threats, making techniques like Random Forest and Decision Trees valuable. Additionally, ethical considerations arise concerning the use of sensitive network data. By addressing these critical issues, the researchers can select appropriate techniques like Random Forest or XGBoost for effective network traffic classification while ensuring alignment with research objectives, data characteristics, and practical constraints.

The research explores the predictive capabilities of machine learning classification models for distinguishing between malicious and benign traffic. It applies six ML algorithms to an IOT based public dataset, TII-SSRC-23 dataset, which are Random Forest, XGBoost, Decision Tree, Multilayer Perceptron (MLP)[13], Extra Trees, and Support Vector Machine (SVM). The model's performance was tested using the parameters accuracy, precision, F1-score, and AUC_ROC score. Accuracy measures the overall correctness of the model's classifications, whereas precision offers information about false positives and false negatives, respectively.[14] The F1-Score combines precision and recall to provide a balanced model performance. Collectively, these features provide a detailed assessment of the model's ability to consistently identify hostile traffic while lowering the risk of false alarms, thus contributing to effective cybersecurity policy.[15]

II. LITERATURE REVIEW

Previously there have been many studies done on the implementation of Machine Learning in malware analysis, such as Sai Tejaswi's research on 'Malicious Attacks Detection Using Machine Learning' focuses on using machine learning techniques, specifically the KMEANS algorithm, to detect malicious attacks, particularly botnet attacks.[16] Traditional methods include manual packet analysis and signature-based procedures. The paper presents a graphical user interface (GUI) for the system, allowing users to upload datasets, apply KMEANS, perform flow integration, feature extraction, normalization, and employ a decision tree method for detection. [16]. The goal is to enhance the detection of botnet attacks in web-based environments using graph-based features and machine learning, offering a promising solution for improved cybersecurity.

Rastogi's study addresses the growing cybercrime threat and the need to detect malicious URLs using machine learning. The authors propose a model that employs various machine learning algorithms, including Random Forest, Support Vector Machine (SVM), Deep Neural Network (DNN), and Convolutional Neural Network (CNN), to identify malicious URLs in datasets containing both malicious and benign URLs.[17] The results show that this strategy is successful for identifying malicious URLs based on their behavior and features. The research also goes into deep learning approaches such as CNN for detecting fraudulent URLs, providing a comprehensive comparison of algorithms for this purpose.[17]

Not just URLs but QR codes have been in malicious streams as well as addressed by Rosario Lourenco's publication which focused on classifying malicious URLs and QR codes used for spreading malware and phishing.[18] It introduces two methods:

Malware URL Categorization: Using machine learning models to classify URLs as dangerous or benign.

Malicious QR Code Classification: Utilizing deep learning techniques, including CNN and ResNet, to detect malicious QR codes.

The research identifies gaps in malicious URL detection, emphasizing the need for improved datasets and model interpretability.[18] Deep learning models, such as CNNs and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:36:59 UTC from IEEE Xplore. Restrictions apply.

ResNet, outperform classical approaches. The paper highlights the difficulties of dealing with malware, as well as the benefits of early identification. While machine learning shows promise, it has limits, such as vulnerability to adversarial assaults. QR code classification models such as CNN, ResNet, and VGG are recommended, depending on dataset size and available resources. The authors achieved great accuracy using several machine learning and deep learning models, highlighting the promise of these techniques in improving security.[18]

Websites play a significant role in phishing and generating malicious network traffic, as emphasized in Sanaa Kaddoura's research on early malicious website detection. The study utilizes supervised machine learning models, including Random Forest, Artificial Neural Network, K-nearest neighbors, and Support Vector Machine, to analyze a dataset containing application layer and network attributes[19], [20],[21]. The experiments yield strong predictive performance, with the Support Vector Machine achieving 92% accuracy in detecting malicious websites. This research is essential for bolstering internet security by effectively identifying malicious websites. [21].Future work could explore the detection of specific cyberattack types and the implementation of hybrid techniques to enhance website classification.

The research focuses on the TI-SSRC dataset which is extracted from a recent paper emphasizing the importance of diverse and robust datasets for enhancing network intrusion detection systems, driven by machine learning.[22] Many existing datasets lack diversity and relevance to current network conditions, limiting intrusion detection effectiveness. To address this, the paper introduces the TII-SSRC-23 dataset with a wide range of traffic types.[22]. The research includes a feature importance analysis and establishes strong baseline performance for both supervised and unsupervised intrusion detection, contributing to the advancement of intrusion detection models in the ever-changing network security landscape.

III. METHODOLOGY

A. Proposed Workflow

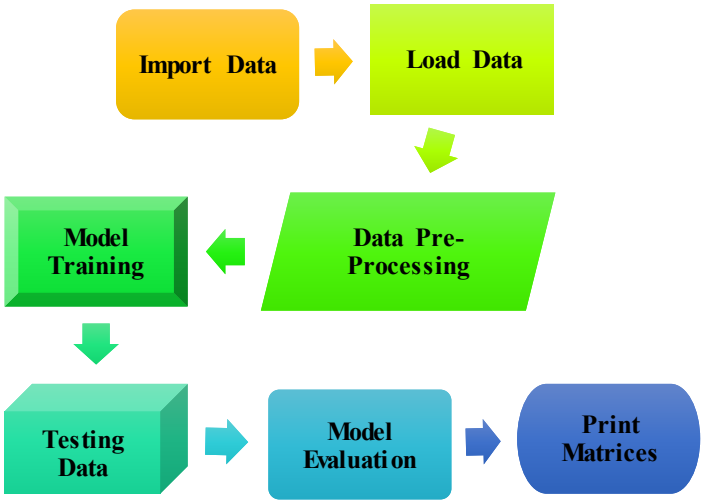


Fig 1: Workflow

The workflow diagram illustrates the research process, beginning with the collection of TII-SSRC-23 network traffic data from Kaggle. Primarily focusing on this data, employing six distinct machine learning models to assess and compare their performance metrics for model efficiency as shown in fig 1.

B. Dataset Used

In this study, the TII-SSRC-23 dataset was used from Kaggle, which contains both benign and malicious network traffic. The benign category has subtypes like audio, background, text, and video, totaling 1300 entries. The malicious category consists of over 1,000,000 datasets, categorized into 4 malicious traffic types with 24 subtypes, including brute force (20%), Dos attack (50%), information gathering (0.04%), and Mirai attack (33%).

Feature Extraction for ML Model:

The dataset contains 1,500,000+ data entries with 81 features. The extracted columns were only 'Label,' 'Traffic Type,' and 'Traffic Subtype' from the network traffic as relevant features. Columns which were excluded were 'Flow ID,' 'Src IP,' 'Src Port,' 'Dst IP,' 'Dst Port,' and 'Timestamp' due to their identifier or time-based nature, which are not relevant for classification. 'Traffic Type' and 'Traffic Subtype' represent network characteristics, while 'Label' is the binary feature indicating benign or malicious. The authors extracted features from each network traffic entry in the training dataset to label them for the ML model. Used a Label Encoder to assign 0 for benign and 1 for malicious labels.

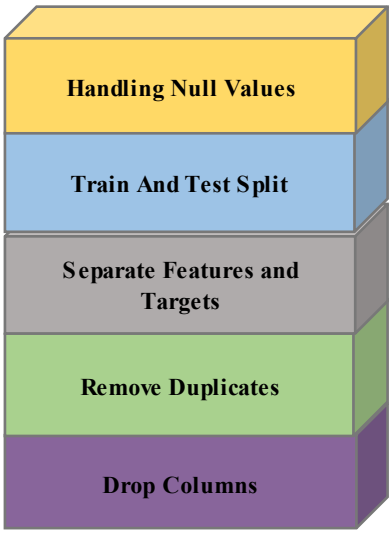
C. Proposed Methodology

1. Import and Load data: Begin by importing essential Python libraries in Jupyter and specifying the columns to drop and the target variables. Then, load the CSV file into the code. Group the data frame by the columns designated as target variables, creating

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:36:59 UTC from IEEE Xplore. Restrictions apply.

separate data frames for each unique occurrence with these target variables.

2. Data Processing:



To prepare the dataset for classification, process the data by removing unnecessary columns, round data to three decimal places, eliminate duplicates, and exclude additional columns based on the target variable. Then, it split the dataset into features (X) and the target variable (y), encode the target variable, and create training and testing sets while preserving the target variable's distribution as shown in fig 2. Duplicates are excluded by feature transformation and preprocessing.

3. Feature transformation and preprocessing: Categorize columns into numerical and categorical groups, even reclassifying the 'Protocol' column. Then, created two preprocessing pipelines: one for numerical columns, handling missing values with mean imputation and standardization, and another for categorical columns, filling missing values with mode and applying one-hot encoding. Using a Column Transformer, streamlined these transformations, fitting it to the training data to learn the process. Finally, apply the preprocessor to both training and testing datasets for further analysis and modeling.
4. Model Training and matrices evaluation:

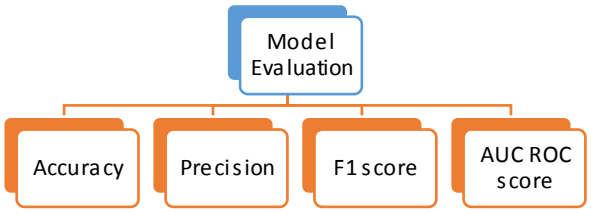


Fig 3: Model Evaluation Matrices

A Random Forest Classifier with 100 trees was utilized, a fixed random state for reproducibility, and optimized CPU core usage for faster training. Following the training, the model was evaluated using sklearn's standard metrics, including accuracy, precision, F1 score, and AUC-ROC score as shown in fig 3.

IV. EXPERIMENTAL ANALYSIS

After analysing the TI-SSRC-23 dataset using various supervised learning models, various calculated evaluation matrices were calculated to measure the performance of each model in classifying the network data as benign or malicious. These matrices are listed below:

- Precision is calculated by:

$$Precision = \frac{T_P}{T_P + F_P} \times 100 \quad -(1)$$

- Recall is calculated by:

$$Recall = \frac{T_P}{T_P + F_N} \times 100 - (2)$$

- F1-Score is calculated by:

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall \times Precision} \times 100 - (3)$$

Higher the F1-score, better the prediction rate. The capacity to discriminate between the various classes is represented by the AUC-ROC measure. The classifier's ability to distinguish between benign and malicious networks is stronger when the AUC-ROC value is higher. This number ought to be less than one and positive.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:36:59 UTC from IEEE Xplore. Restrictions apply.

TABLE I. PERFORMANCE MEASURES

Models	**Accuracy (%)**	**Precision (%)**	**F1-Score**	**AUC-ROC Score**
SVM	99.9	97.1	0.97	0.97
Random Forest	99.7	93.0	0.93	0.99
Decision Tree	99.7	88.7	0.88	0.94
Extra Tree	99.7	93.3	0.93	0.98
MLP	99.7	79.1	0.77	0.98
XGBoost	99.8	93.6	0.92	0.99

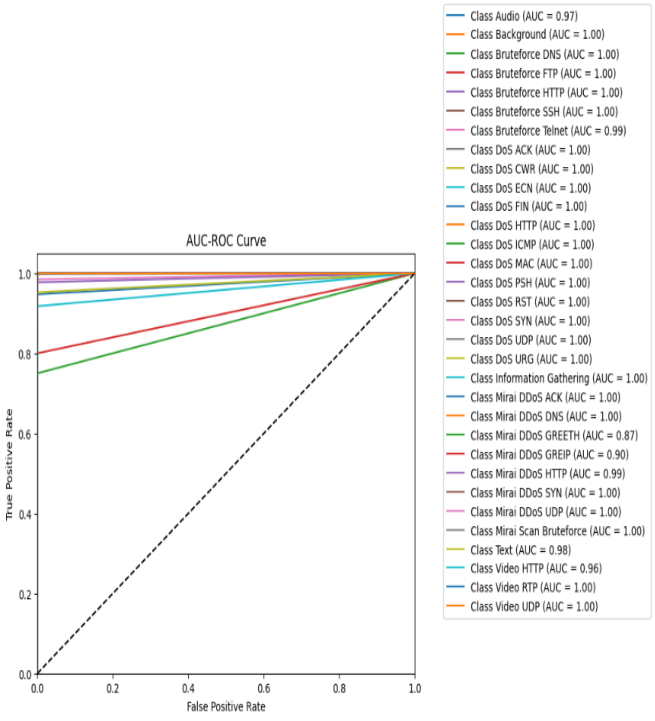


Fig 4: AUC-ROC Curve for Random Forest

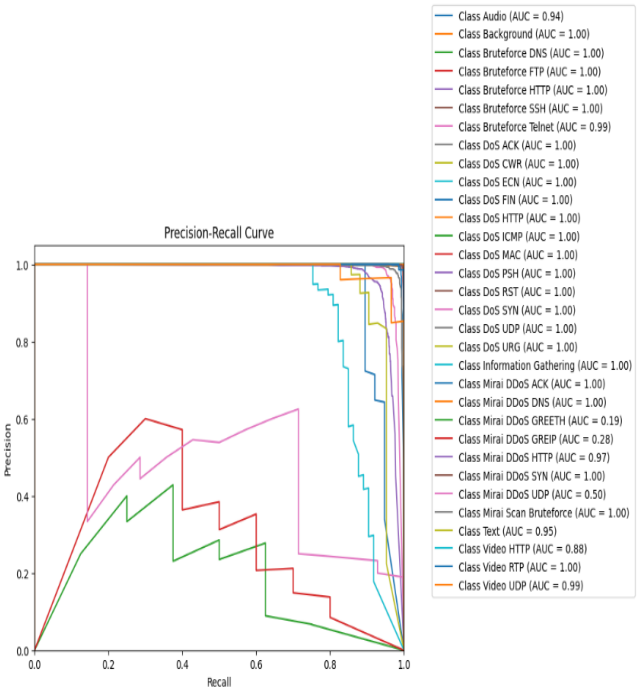


Fig 5: Precision-Recall Curve for Random Forest

The hardware requirements for executing these models are Google Collab Pro with a K80 GPU, 2 x vCPU CPU, 32GB RAM, and a Tesla T4. This approach ensures that the models are processed efficiently and accurately. These specifications are ideal for running such large datasets without encountering any software issues.

V. RESULTS AND DISCUSSION

Using the TII-SSRC-23 dataset, the study compares the effectiveness of six machine learning models in categorizing network traffic as benign or malicious. Random Forest, Decision Tree, Extra Tree, Multilayer Perceptron (MLP), XGBoost, and Support Vector Machine (SVM) were the models tested. The evaluation measures include accuracy, precision, F1-score, and AUC-ROC score, and the analysis provides insight into the performance of each model. The findings show that all models were highly accurate, with the following values:

a) Random Forest: 99.7%
b) Decision Tree: 99.7%
c) Extra Tree: 99.7%
d) MLP: 99.7%
e) XGBoost: 99.8%
f) Support Vector Machine (SVM): 99.9%

As a result, it was determined that the Support Vector Machine had the highest accuracy at 99.9%. Precision, which assesses the true positive rate among anticipated positives, shows that the Support Vector Machine had 97.1%. The F1-score, which balances precision and recall, ranks Support Vector Machine and Random Forest as the top performers,

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:36:59 UTC from IEEE Xplore. Restrictions apply.

with 0.97 and 0.93, respectively. The AUC-ROC score, which measures the classifier's ability to distinguish between benign and malicious networks, identifies Random Forest and XGBoost as the most robust models, with both earning 0.99. As a result, since Random Forest and XGBoost have the largest AUC-ROC curves, they are both the most effective classifiers for determining whether a network is benign or malevolent. By further comparing results it can be concluded that Random Forst is slightly better than XGBoost as it has a higher F1-score. The Precision-Recall curve, fig 5, and AUC-ROC curve, fig4, for Random Forest illustrate the model's excellent performance in distinguishing between benign and malicious instances. Along with the accuracy it is also important to calculate the classification error which is defined as the compliment of accuracy. The classification error is calculated by subtracting accuracy from 100%. Here, random forest has the accuracy which is 99.7%. Hence its classification error would be 100-99.7=0.3%. SVM has the least classification error with only 0.1%.

The experimental analysis suggests that Random Forest and XGBoost are highly effective in classifying network traffic as benign or malicious based on the TII-SSRC-23 dataset. These models demonstrate robustness in accuracy, precision, and overall performance. However, the paper also acknowledges the potential of neural networks and suggests future work that explores the use of the latest classifiers and hybrid approaches for improved network categorization. The focus on recognizing specific cyberattack types and identifying harmful behavior in incoming network traffic indicates a forward-looking approach to enhancing cybersecurity measures. Overall, the study contributes valuable insights into the application of machine learning models for network security in the evolving landscape of cyber threats.

VI. Conclusion and Future Work

The research has applied six different supervised machine learning models to binary classify the incoming network dataset as either benign or malicious on the publicly available dataset of TI-SSRC-23. All these models were trained on the same dataset by which their accuracy, precision, F1-score, and AUC-ROC scores have been evaluated and analysed. The experimental analysis for classifying the network as benign or malicious showed that the highest F1-score metric among all the algorithms was for support vector machine with 0.97. But when compared with AUC-ROC score, Random Forest and XGBoost had the highest score above all other algorithms with 0.99. Overall, on average all performance matrices, Random Forest showed the best results for predicting the outcome therefore the model loss curve and training curve was plotted.

Since this research used the state-of-art machine learning models, for future work it would be more efficient to use neural network and latest classifiers to predict the network traffics. Not only is it important to identify dangerous networks, but it's also critical to identify the harmful behaviour of the incoming network traffic in order to deliver the exact countermeasure in the shortest amount of time. Future research will take into account creating a system that can recognize and identify the kind of cyberattack that a person could experience. Secondly, employing hybrid approaches provides an additional avenue for improving network categorization.

References

[1] K. Khalil, K. Elgazzar, M. Seliem, and M. Bayoumi, "Resource discovery techniques in the internet of things: A review," *Internet Things*, vol. 12, p. 100293, Dec. 2020, doi: 10.1016/j.iot.2020.100293.

[2] O. P. Suman, "A Novel Approach for Malicious Domain Classification Based on Dns Traffic Analysis and Machine Learning." Rochester, NY, Oct. 04, 2023. doi: 10.2139/ssrn.4592811.

[3] A. Almuqren, M. Frikha, and A. Albuali, "Automated Malware Detection Based on a Machine Learning Algorithm," in *2023 IEEE Tenth International Conference on Communications and Networking (ComNet)*, Nov. 2023, pp. 1–12. doi: 10.1109/ComNet60156.2023.10366550.

[4] "1.4. Support Vector Machines," scikit-learn. Accessed: Nov. 07, 2023. [Online]. Available: https://scikit-learn/stable/modules/svm.html

[5] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Their Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998, doi: 10.1109/5254.708428.

[6] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[7] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.

[8] A. Verma, V. Ranga, and D. K. Vishwakarma, "A novel approach for forecasting PM2.5 pollution in Delhi using CATALYST," *Environ. Monit. Assess.*, vol. 195, no. 12, p. 1457, Nov. 2023, doi: 10.1007/s10661-023-12020-z.

[9] K. Tripathi and A. Verma, "Remote Sensing and Machine Learning Fusion: A Robust Framework for Land Use and Land Cover Change Detection," in *2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE)*, IEEE, 2024, pp. 434–439. Accessed: May 09, 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10503360/

[10] G. Singh and M. Sachan, "Multi-layer perceptron (MLP) neural network technique for offline handwritten Gurmukhi character recognition," in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, Dec. 2014, pp. 1–5. doi: 10.1109/ICCIC.2014.7238334.

[11] L. Rokach and O. Maimon, "Decision Trees," in *The Data Mining and Knowledge Discovery Handbook*, vol. 6, 2005, pp. 165–192. doi: 10.1007/0-387-25465-X_9.

[12] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely Randomized Trees," *Mach. Learn.*, vol. 63, pp. 3–42, Apr. 2006, doi: 10.1007/s10994-006-6226-1.

[13] A. Verma, V. Ranga, and D. K. Vishwakarma, "Forecasting of Satellite Based Carbon-Monoxide Time-Series Data Using a Deep Learning Approach," in *2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Feb. 2023, pp. 1–4. doi: 10.1109/ICITIIT57246.2023.10068609.

[14] "Malware Family Classification using Active Learning by Learning | IEEE Conference Publication | IEEE Xplore." Accessed: Jan. 13, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/9061419

[15] S. Choudhary and A. Sharma, "Malware Detection & Classification using Machine Learning," in *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, Feb. 2020, pp. 1–4. doi: 10.1109/ICONC345789.2020.9117547.

[16] C. S. Tejaswi, Y. Chaitanya, A. Jesudoss, and P. Shyry, "Malicious Attacks Detection Using Machine Learning," in *2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA)*, Sep. 2022, pp. 924–929. doi: 10.1109/ICIRCA54612.2022.9985551.

[17] P. Rastogi, E. Singh, V. Malik, A. Gupta, and S. Vijh, "Detection of Malicious Cyber Fraud using Machine Learning Techniques," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2022, pp. 520–524. doi: 10.1109/Confluence52989.2022.9734181.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:36:59 UTC from IEEE Xplore. Restrictions apply.

[18] D. O. Do Rosario Lourenco, M. V. H. Sai Sriraj, K. K. Thambi, and V. Ranjan, "Malicious URLs and QR Code Classification Using Machine Learning and Deep Learning Techniques," in *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, Aug. 2023, pp. 1–10. doi: 10.1109/ASIANCON58793.2023.10270125.

[19] A. Verma, V. Ranga, and D. K. Vishwakarma, "Combating Respiratory Health Issues with Intelligent NO2 Level Prediction from Sentinel 5P Satellite," in *2023 IEEE 20th India Council International Conference (INDICON)*, Hyderabad, India: IEEE, Dec. 2023, pp. 882–886. doi: 10.1109/INDICON59947.2023.10440910.

[20] A. Verma, V. Ranga, and D. K. Vishwakarma, "BREATH-Net: a novel deep learning framework for NO2 prediction using bi-directional encoder with transformer," *Environ. Monit. Assess.*, vol. 196, no. 4, p. 340, Apr. 2024, doi: 10.1007/s10661-024-12455-y.

[21] S. Kaddoura, "Classification of malicious and benign websites by network features using supervised machine learning algorithms," in *2021 5th Cyber Security in Networking Conference (CSNet)*, Oct. 2021, pp. 36–40. doi: 10.1109/CSNet52717.2021.9614273.

[22] D. Herzalla, W. T. Lunardi, and M. Andreoni, "TII-SSRC-23 Dataset: Typological Exploration of Diverse Traffic Patterns for Intrusion Detection," *IEEE Access*, vol. 11, pp. 118577–118594, 2023, doi: 10.1109/ACCESS.2023.3319213.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:36:59 UTC from IEEE Xplore. Restrictions apply.

J Mater Sci

Composites & nanocomposites

A flexible memory device made of SnO_2-hBN nanocomposite exhibits stable resistive switching application

Km. Komal[1], Mukhtiyar Singh[1], and Bharti Singh[1,*]

[1] *Department of Applied Physics, Delhi Technological University, Delhi 110042, India*

Received: 28 March 2024
Accepted: 2 July 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature, 2024

ABSTRACT

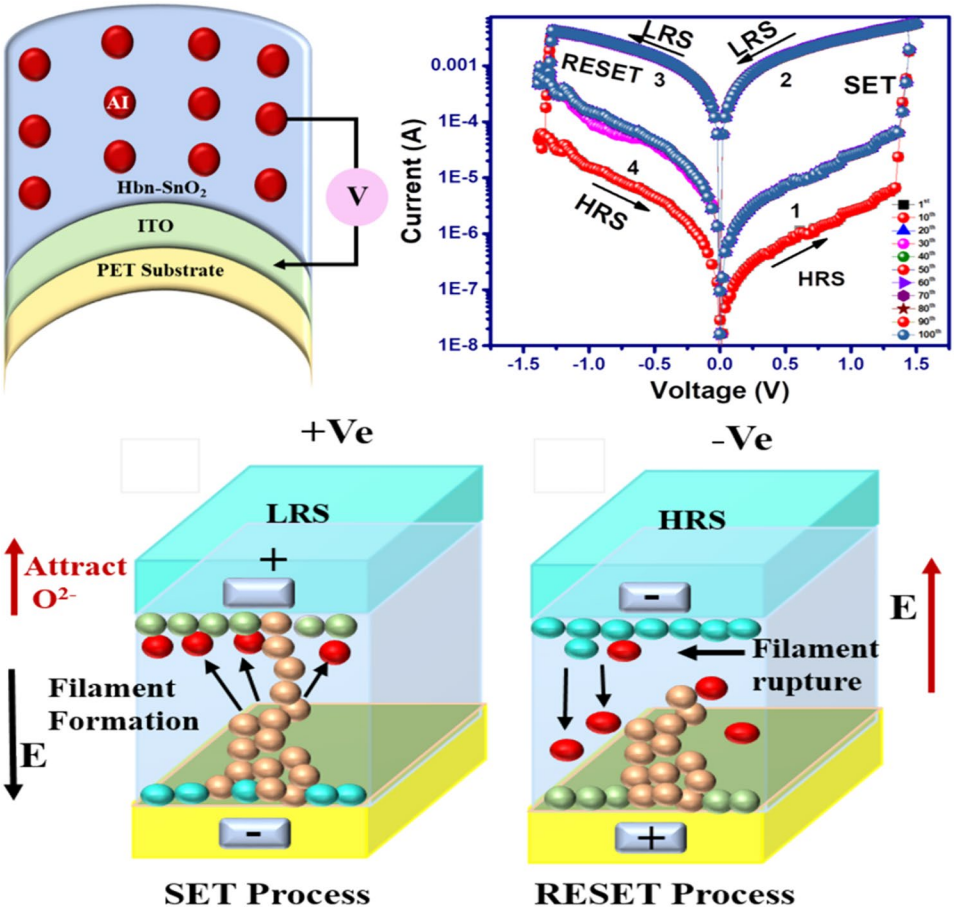
Recent developments in flexible memories, neuromorphic computing, and other fields have generated a lot of interest in memristive devices based on two-dimensional materials. Due to its capacity for data storage and quick working speed, two-terminal non-volatile memory devices are more efficient and incredible. The need for memory that is flexible, low-powered, and ultra-thin is rising as a result of the rapidly developing fields of healthcare, wearable electronics, and the Internet of Things. The current study observes the effects of pure hexagonal boron nitride (hBN) inclusion on the resistive switching properties of tin oxide (SnO_2)-based flexible resistive memory device, which was synthesized by a straightforward hydrothermal method. The synthesized nanocomposite powders were characterized by X-ray diffraction and Raman spectroscopic methods. A pure SnO_2 and SnO_2-hBN (5, 7, and 10 wt.%) nanocomposite film was spin-coated on an ITO-PET flexible substrate to create the memory device. To finish the memory device, thermal evaporation was used to create Al top electrodes with a thickness of less than 100 nm. After comparing the resistive switching performance of the created devices, it was found that the memory device using 7% hBN in SnO_2 had a higher I_{ON}/I_{OFF} ratio of 1000, whereas the samples including 5 and 10 wt.% hBN had ratios of 96 and 10, respectively. In addition, the retention and endurance behaviour of the fabricated 7%SnO_2-hBN composite film-based device was studied for up to 1.5×10^4 and 100 cycles, respectively, with no degradation in the memory window. It has been noted that the current metal oxide–2D hybrid nanocomposite is a promising option for low-power, flexible, and enduring non-volatile memory applications.

Handling Editor: Peiyao Zhao.

Address correspondence to E-mail: bhartisingh@dtu.ac.in

https://doi.org/10.1007/s10853-024-09976-9
Published online: 16 July 2024

GRAPHICAL ABSTRACT



Introduction

To address the issue of huge information storage and memory scalability, both the industrial and academic groups are working to develop innovative non-volatile memory technologies to replace the current flash memory [1]. The fact, that flexible devices can be bent, stretched, and used in wearable electronics, has undoubtedly drawn a lot of interest in these gadgets [2]. Recently, there has been a growth in the requirement for flexible, wearable, foldable, lightweight, and transparent electronic devices in many aspects of modern life. Intelligent devices are gaining popularity as a result of the introduction of 5G communication networks, and nonlinear electric transport materials are now seen as a significant boost for smart devices [3]. Examples of these devices include mobile phones, radio frequency identification tags, sensors, displays, and watches. The essential component of intelligent flexible electronics is the data processing and storage systems. In order to address the demands of wearable gadgets, researchers are dedicated to creating flexible mersistors [4]. An increasing number of people are becoming interested in novel non-volatile memory devices with bright futures, such as resistive random access memory (RRAM), ferroelectric random access memory (FeRAM), magnetoresistive random access memory (MRAM), and phase-change random access memory (PCRAM) [5]. Recent research has focused more and more on RRAM due to its higher memory performance, and it is expected to be one of the most promising options for next-generation memory. However, the small memory window, precise material and structure, and other reasons limit MRAM. Because of this, there are still several problems in the way of applying the three previously mentioned types of memories in large-scale memory [6]. The

RRAM device, which can be reduced down to less than 10 nm and provides high write/erase speed, high durability, and low-power consumption, has drawn great interest and has the potential to substitute charge-based memory devices [7]. A wide range of materials showing resistive switching behaviours, including metal oxides, organic polymers, amino acids, chalcogenides, and 2D nanotechnologies, have been studied as potential candidates for memory devices [8]. Hickmott originally reported the resistive switching phenomenon in 1962. This was confirmed when Hickmott found the hysteretic current–voltage (I–V) property in a metal–insulator–metal (MIM) structure [1]. Thus, the 1970s and 1980s marked the beginning of a period of intense research into the resistive switching phenomenon. Early studies primarily addressed and elucidated the physical mechanism of electrically stimulated resistive switching. The second resistive switching research surge began in the late 1990s when scientists realized that resistive switching behaviour has the potential to be used as an ultimate NVM as microelectronics processing technology developed [9]. Metal–insulator–metal (MIM) capacitor structures are a usual form of resistive switching (RS) device. By putting a device under electrical stress, it is possible to switch its resistance state from high resistance (HRS) to low resistance (LRS). This RS behaviour is caused by the assembly of metal ions moving away from a working metal electrode or donor-type point defects, which results in the creation and destruction of conducting filaments (CFs) in the dielectric [10]. To create a high-performance RRAM device, the RS phenomenon has been investigated in several material categories. The main contributors include low-dimensional materials such as graphene oxide (GO), molybdenum disulphide (MoS_2), reduced GO, tungsten disulphide, and hexagonal boron nitride; polymer dielectrics such as poly(4-vinylphenol) (PVP), polyvinyl carbazole, polymethyl methacrylate, and polyvinyl alcohol; and metal oxides such as HfOx, TiOx, AlOx, TaOx, and ZnO; these materials are among the main contributors [11]. In recent times, hybrid RRAMs have gained significant attention, mainly due to their reliable and low-power switching operations [12]. The term hybrid RRAM is attributed to the nonconventional multilayer structure of the device with either organic polymer composites or metal oxide or both, where the metal oxide layer enhances the memory window, and the polymer composite contributes towards improved reliability [13]. However, these devices high ON-current remains an important problem that needs to be fixed to increase their power efficiency.

In recent times, two-dimensional (2D) materials have gained tremendous attention and growth owing to their excellent performance in applications not limited to field effect transistors, sensors, photodetectors, non-volatile memories, and solar cells [14]. Hexagonal boron nitride (hBN) is an intriguing ultra-wide bandgap insulating material with a crystal structure resembling graphite [15], where boron and nitrogen atoms alternately replace the sp^2-hybridized carbon atoms. Because of its graphite-like structure, hBN has great chemical and thermal stability and strong thermal conductivity, making it a viable choice for oxidation resistance and heat spreading [16]. Additionally, hBN provides a notable improvement in the confinement effects of chemical and physical characteristics, as well as a high number of active sites. Very few studies have been reported on hexagonal boron nitride (hBN), despite most of the studies on the resistive switching effect in 2D materials primarily focuses on graphene and molybdenum disulphide (MoS_2). Flexible resistive switching applications for hBN and its nanocomposites still have a lot of unexplored potential. These hybrid nanocomposites are expected to have a configurable bandgap, a high level of flexibility, and a combination of unusual electrical and mechanical properties as a result more research is required to extract more insightful conclusions [17]. Additionally, Table 1 compares the findings from our study with those from other published research.

SnO_2 was chosen as a switching material because it is an n-type semiconductor that is conducted by oxygen vacancies and was reported to have high thermal stability [26]. In the literature, SnO_2-based sensors have demonstrated good endurance. Because the Frenkel defect energy (7 eV) is substantially greater than the bandgap (3.6 eV), SnO_2 is resistant to displacement damage. Because of these features, SnO_2 is an excellent contender for resistive memory applications in radiation-hard situations. However, further research is required to completely comprehend the MIM devices filament and current mechanisms and to increase its reliability. For this reason, SnO_2 has received limited attention even though it has shown promising properties for resistive switching [27]. SnO_2 is an n-type broadband semiconductor in its purest form, and the electrical conductivity of

Table 1 Table showing the comparison of our devices with other RRAM devices

Device	$V_{set(V)}$	$V_{reset(V)}$	Operation mode	I_{ON}/I_{OFF} ratio	Endurance	Retention	References
Ag/SnO_2/FTO	−15.7	+17	Bipolar	–	10 cycles	–	[18]
Al/Sb:SnO_2/TiO_2/Al	−3.0	+1.5	Bipolar	9×10^1	40 cycles	–	[19]
Al/RGO-SnO_2/ITO	+4.93	−3.21	Bipolar	4.11	100 cycles	$1*10^3$s	[20]
SnO_2	+20	−20	Bipolar	–	–	–	[21]
Ag/MoS_2-SnO_2/quartz substrate	+5	−5	Linear behaviour	–	–	–	[22]
Ag/SnO_2:Fe/Ag	+2	−2	Bipolar	18	–	–	[23]
Al/MoS_2-SnO_2/ITO-PET	+1.96	−1.52	Bipolar	25	–	–	[24]
Al/MoS_2-SnO_2/ITO-PET	+0.89	−0.85	Bipolar	100	100	2×10^3 s	[25]
Al/hBN-SnO_2/ITO-PET	+0.5 V	−0.7 V	Bipolar	1000	100	1.5×10^4 s	This work

this material is caused by structural flaws or, more specifically, departures from stoichiometry in the lattice. It has been demonstrated that oxygen vacancies in the SnO_2 lattice are the predominant structural flaws in pure SnO_2 thin films, suggesting that tin atoms across the lattice may exist on a microscopic level as well. The electrically conductive properties of transparent SnO_2 semiconductors thin layers can be increased by adding donor impurities. In thin layers of SnO_2, the amount of free electrons is decreased upon the addition of impurities with a valence of less than four, such as zinc (Zn), which lowers the thin film's electrical conductivity. However, the converse happens when impurities like Sb, which have a valence of more than four, are added; as a result, thin films conduct electricity more readily. Antimony (Sb) and fluorine (F) are common contaminants for SnO_2 thin films. The thin coating's electrical conductivity becomes more thermally stable when Sb is added to SnO_2. Also, the electrical conductivity of transparent SnO_2 semiconductor thin films can be tailored by adding hexagonal boron nitride (hBN) in its matrix [28].

In the current study, pure SnO_2 and SnO_2-hBN with the different wt.% of hBN nanocomposite powder was made using the simple hydrothermal process [22] and was used further for spin coating the active memory layer in MIM configuration. The resistive switching performance of devices based on Al/SnO_2/ITO-PET and Al/SnO_2-hBN/ITO-PET was observed in relation to hysteresis loop, durability, and retention tests. The logarithmic plot of I–V characteristics allowed for a comprehensive analysis of the electrical charge transport mechanism responsible for the resistive switching behaviour. The ability to get constant resistive switching outcomes even after bending down 100 times was used to illustrate the flexibility of the ReRAM devices that were created.

Experimental

Substances

The ingredients used to make tin (IV) chloride pentahydrate ($SnCl_4.5H_2O$) are as follows: 98% of the boron nitride powder (BN) comes from Sigma-Aldrich. Sodium hydroxide (NaOH, 98%), ethanol (C_2H_5OH, 99.9%), isopropanol (C_3H_7OH, IPA), and hydrazine hydrate ($N_2H_4H_2O$, > 99%) are purchased from Fisher Scientific. Since the compounds were in analytic purity, no more purification was necessary.

Synthesis of hBN flakes

A simple in situ hydrothermal process has been used to produce hexagonal boron nitride nanosheets. The dispersion is typically made by mixing 50 ml of isopropyl alcohol (C_3H_7OH) and 10 ml of hydrazine hydrate ($N_2H_4.H_2O$) with 0.75 g of bulk hBN powder. Add 50 ml of deionized (DI) water to this dispersion and mix for 2 h. The final step is to transfer the resulting dispersion into a 100-ml Teflon-lined stainless steel autoclave vessel and seal it. An oven with a high temperature is used to keep the autoclave at 220 °C for 20 h. The bulk BN powder nanosheets were exfoliated in a stainless steel autoclave. Ultimately, the finished product was dried for 12 h at 90 °C after being repeatedly cleaned with DI water. The white powder in Eppendorf is gathered and kept until it is required at room temperature (RT-25 °C).

hBN-SnO_2 nanocomposite synthesis

To synthesize 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposites, firstly, synthesized hBN (200 mg, 280 mg, and 400 mg) was dissolved in 4 ml of DI water, then ultrasonically sonicated for 30 min. Furthermore, $SnCl_4{\cdot}5H_2O$ (1.76 g) and NaOH (2 g) were dissolved in a mixture of ethylene glycol (12 ml) added into 88 ml DI water and magnetically stirred for 1 h to get a homogeneous solution [29]. The SnO_2 solution was then added to the hBN resolution after 1 h of stirring and stirred for 30 min. After the suspension was prepared, it was placed inside a 100-mL Teflon-lined stainless steel autoclave and heated to 180 °C for 18 h. Both ethanol and DI water were used for three centrifugations of the material. The leftovers were dried for 5 h at 50 °C in a vacuum oven [30]. The two-step hydrothermal process for making SnO_2-hBN nanocomposite is depicted in Fig. 1.

Fabrication of films

On top of ITO-PET substrates, a thin layer of SnO_2-hBN nanocomposite was placed. First, a normal wet cleaning process was used to clean the ITO-PET substrates. After washing, the materials were dried by heating them to 100 °C for 20 min. The solution was created by ultrasonically mixing 100 mg of SnO_2-hBN powder with 3 ml of ethanol solution for 45 min, then applying the mixture to the ITO-PET substrate (2 cm × 1 cm) and spinning it for 60 s at 2000 rpm. The sample covered in coatings was placed on a heated plate for 10 min at 100 °C. After seven cycles of spin coating, a homogeneous layer of SnO_2-hBN was formed. Finally, the top aluminium electrodes were deposited through the shadow mask using the thermal evaporation process under vacuum conditions of 10^{-5} bar [20]. The fabrication process of the Al/SnO_2-hBN/ITO-PET-based MIM device is schematically shown in Fig. 1 [31].

Materials characterization techniques

First, made-up samples are examined with the WiTec alpha 300RA Raman Spectrometer, which uses a laser source with a wavelength of 530 nm. An X-ray diffractometer (Bruker D8 Advance) and a Cu Kα source (λ = 1.54 Å) were used to study the structural phase of hBN, SnO_2, and SnO_2-hBN (5%, 7%, and 10%) materials. The hBN, SnO_2, and SnO_2-hBN nanocomposite materials surface morphologies were investigated using electron microscopy (FESEM; ZEISS). Using a monochromatic Al Kα source and a Perkin Elmer ultra-high vacuum XPS machine (Model 1257 PHI), the interfacial analysis of hBN, SnO_2, and SnO_2-hBN nanocomposite structures was achieved. The thin layer of synthesized SnO_2-hBN nanocomposite film is deposited using the spin coating process. Aluminium electrodes less than 100 nm were made for memory devices using the thermal evaporation technique. After that, I–V measurements were performed on the created thin-film memory cells (Al/SnO_2/ITO, Al/hBN-SnO_2/ITO) using a Keithley Source Metre 2450. First,

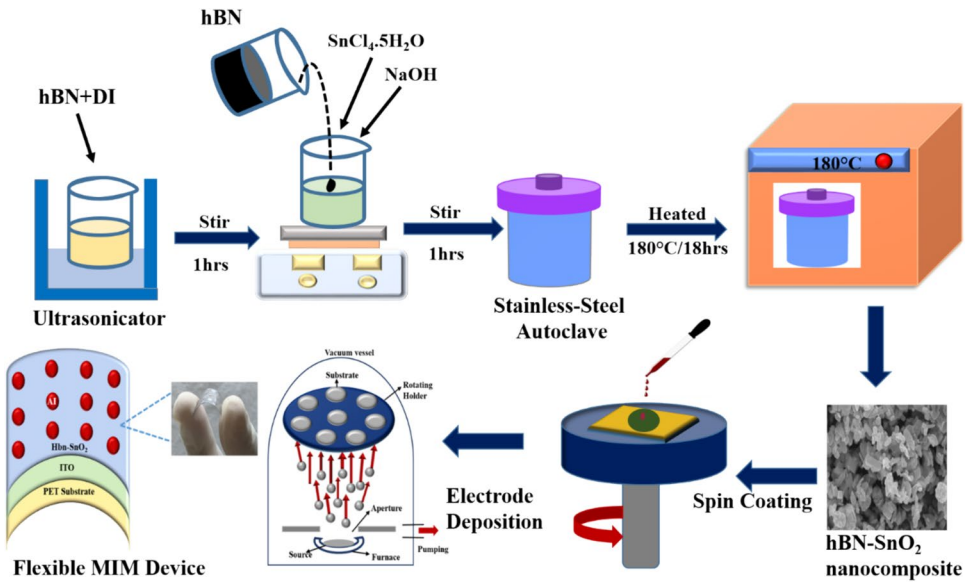


Figure 1 A visual representation showing the synthesis of hBN-SnO_2 (5%, 7%, and 10%) nanocomposite powder using the hydrothermal method and the pictorial presentation of steps involved in the fabrication of resistive memory devices.

we tested the dielectric using the capacitance method on a sample that had been heated to 260 °C for 2 h and covered with silver paste on both sides. With an Agilent LCR metre (Model E4980A), the frequency (20–10^6 Hz) and loss tangent ($D = \tan\delta$) measurements were compared to one another.

Results and discussion

X-ray diffraction

The XRD technique was used to examine the crystal structure of synthesized pure hBN, pure SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposite powder. As illustrated in Fig. 2a, XRD spectra verify that hBN and SnO_2 exist in the SnO_2-hBN nanocomposite. The existence of tetragonal rutile structure in pure SnO_2 is confirmed by its XRD spectra, which show peaks at 26°, 34°, 52°, and 65°, corresponding to the (110), (101), (211), and (202) planes. Similar to this, the as-synthesized hBN samples peaks are indexed at 26.86°, 41.8°, 44.01°, 50.5°, 55.34°, and 76.1°, which correspond to the hexagonal boron nitride (002), (100), (101), (102), (004), and (110) distinctive hkl planes. The XRD data showed the development of a pure hexagonal phase of BN with good crystallinity, with no extra impurity peaks detected. In comparison with the diffraction planes (100), (101), (102), (004), and (110), the intensity of the peak connected to the (002) diffraction plane shown in Fig. 2a is significantly higher. This suggests that the hBN (002) crystal plane is more exposed. The tetragonal rutile phase of SnO_2 is further confirmed for the SnO_2-hBN nanocomposite, which displays peaks at 26°, 34°, 52°, and 65° that are related to the (110), (101), (211), and (202) planes.

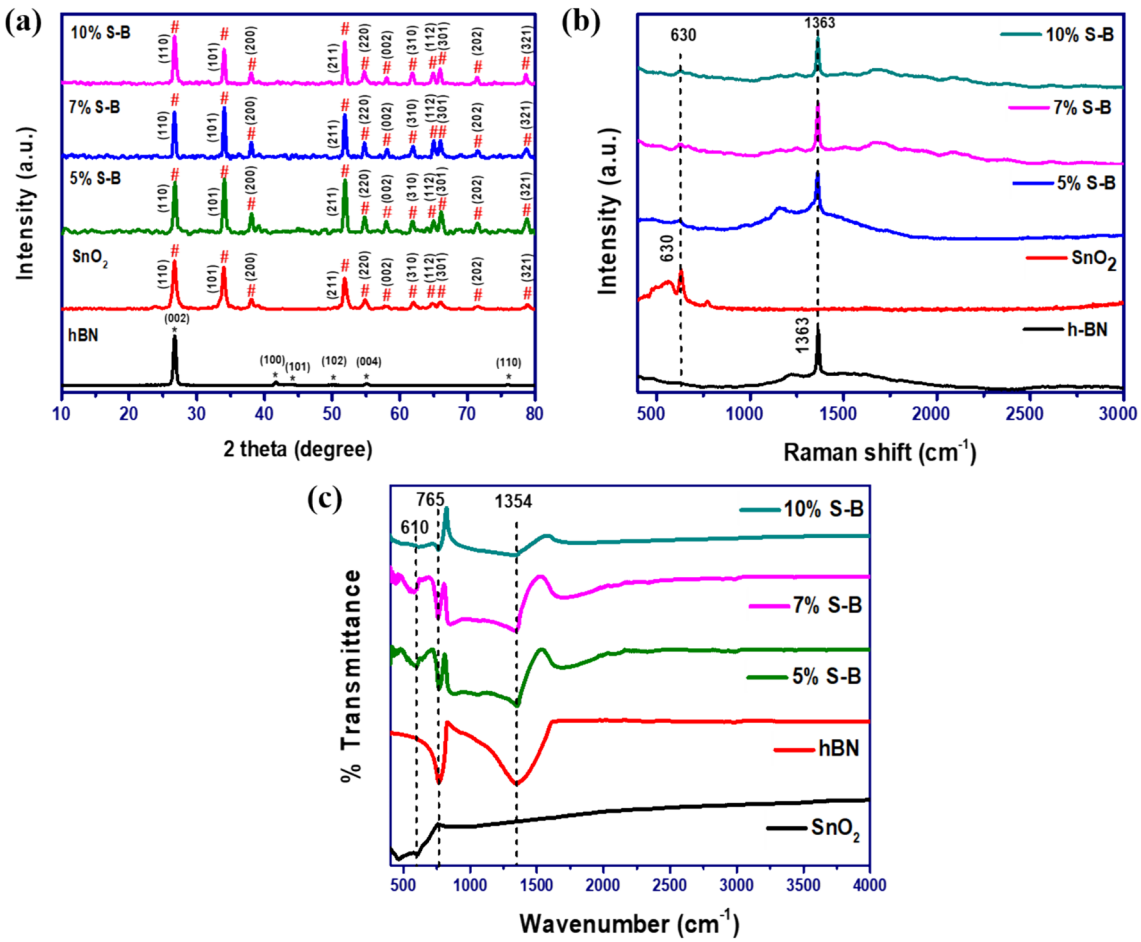


Figure 2 **a** XRD diffractogram, **b** Raman spectroscopy, and **c** FTIR analysis of hBN, SnO_2, and hBN-SnO_2 nanocomposite with the different weight percentages of hBN powder.

Because hBN has a low crystallinity, the peaks of pristine hBN are not visible in the SnO_2-hBN nanocomposite. The pristine SnO_2 diffraction peak, which corresponds to the (110) plane, dominates the hBN peak in the composite sample, which is visible at 26° (002). It could be explained by the SnO_2-hBN samples transformation of Sn^{4+} into SnO_2 nanoparticles [32]. The addition of hBN was found to cause the SnO_2 diffraction peaks to expand. As crystallinity depends on crystallite size, the crystallite size was calculated by using the Scherrer formula for pure hBN, pure SnO_2, and SnO_2-hBN nanocomposite [20, 33];

$$D = k\lambda/(\beta\cos\theta) \quad (1)$$

where λ denotes the wavelength of X-ray, β denotes the broadening of the diffraction peak measured at half maximum intensity (FWHM), θ denotes the angle corresponding to peak position, crystallite size is D, and k denotes the shape factor ($k = 0.94$).

The average crystal size determined for pure SnO_2, pure hBN, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposites are calculated to be 13.2 nm, 19.5 nm, 26.1 nm, 20.8 nm, and 20.7 nm, respectively. It has been observed that the average crystallite size of SnO_2-hBN nanocomposite decreases owing to the growth and nucleation of SnO_2 crystals, which was hindered by hBN [34, 35].

Raman characterization

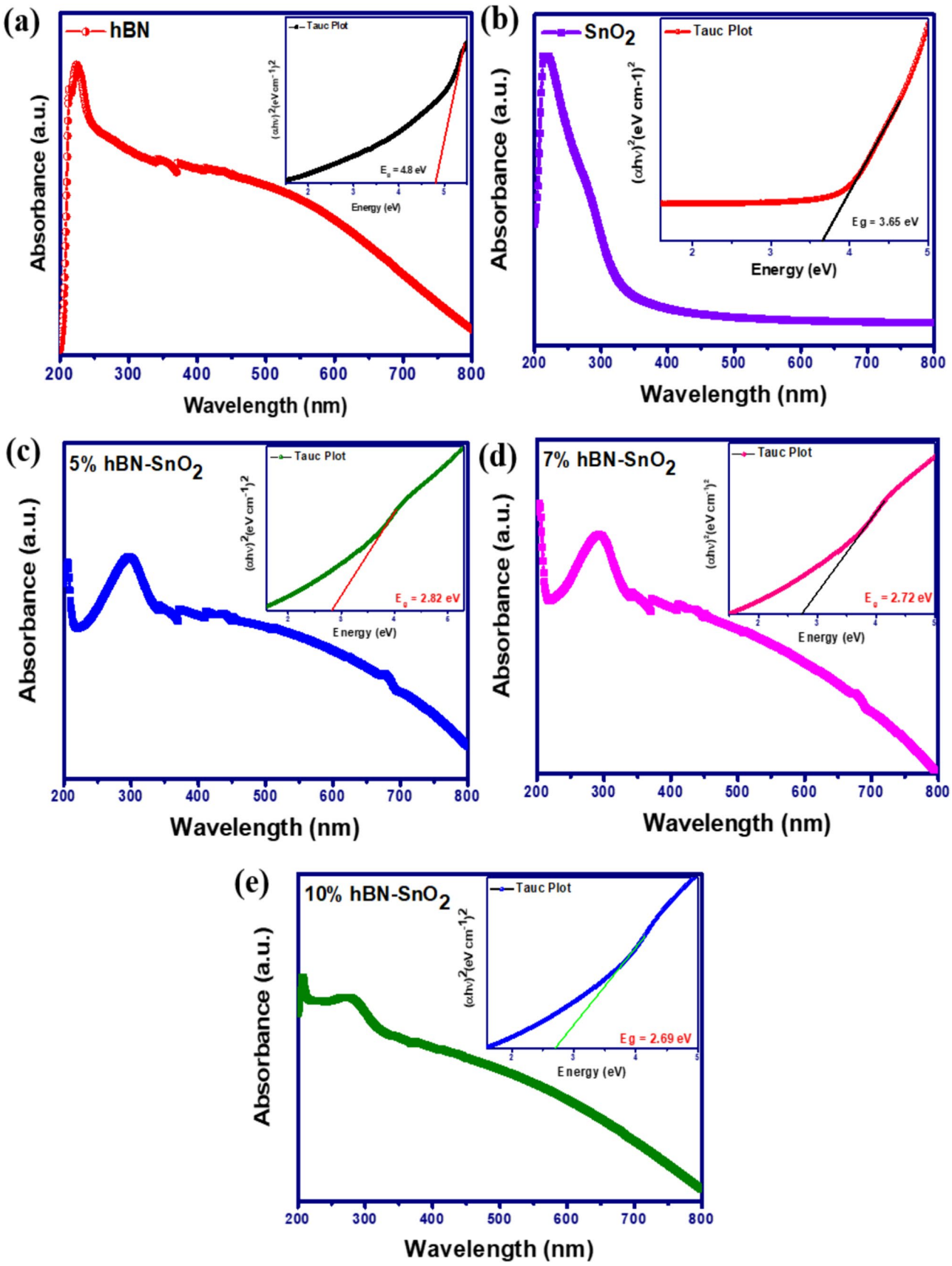
As Raman scattering is an effective process for examining the structure of hBN-based substances, it was used to investigate the SnO_2-hBN nanocomposite structure in more detail. Figure 2b displays the Raman spectra of SnO_2-hBN nanocomposites and SnO_2 nanoparticles at different hBN concentrations. The Raman peak of SnO_2 nanoparticles at the peak position of 630 cm^{-1} corresponds to the A_{1g} vibration mode [36], whereas, for the bulk hBN, the characteristics peak is observed at 1363 cm^{-1}, corresponding to in-plane vibrational mode (E_{2g}), respectively [15, 17]. Raman spectra for SnO_2-hBN nanocomposite powder reveal distinct peaks at different hBN concentrations (5, 7, and 10%). For the 5%SnO_2-hBN nanocomposite sample, the Raman spectrum is a simple superposition of the single vibration modes of hBN and SnO_2, which is consistent with the literature [37]. Peak positions remain unchanged, but the hBN vibration modes E_{2g} and A_{1g} have significantly lower peak intensities. As the hBN content increases, the peak intensity of the SnO_2 A_{1g} mode decreases, signifying the successful creation of the nanocomposite material with a variable hBN content.

FTIR spectra

The chemical interactions and advantageous groups present in a substance can be further investigated using the FTIR method. Since it shows the composition of the material, knowing which functional groups are adsorbed at certain occurrences under the transmission method is very useful. Band locations and the number of absorption peaks are determined by morphology, chemical composition, and crystalline structure [38]. The surface of nanoscale materials is extremely sensitive and easily altered. An FTIR study at room temperature was performed to examine the compound structures on the outer layer of the materials over a wavenumber between 400 and 4000 cm^{-1}. Figure 2c displays the FTIR spectra of pure hBN, pure SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposites. Pure hBN exhibits two strong peaks in its FTIR spectra at 790 and 1378 cm^{-1}, respectively. The broad B-N stretching was observed, including in-plane stretching vibration (B-N, 1378 cm^{-1}) and the out-of-plane bending mode (B-N, 809 cm^{-1}), as shown in Fig. 2c [39]. The band exhibited in the low wavenumber region 610 cm^{-1} is attributed to the Sn–O stretching vibration [38, 40]. Characteristic peaks of SnO_2-hBN occurred at 1378 cm^{-1}, 709 cm^{-1} (B-N stretching vibration), and 610 cm^{-1} (Sn–O vibration); furthermore, the peak positions do not change; however, the peak intensities of the hBN vibration modes B-N are dramatically reduced and also the peaks are broadened. With the increase in hBN content, the peak intensity for the Sn–O mode of SnO_2 decreases, indicating the successful fabrication of the nanocomposite substance with variable hBN content [41].

Ultraviolet–visible spectroscopy

In Fig. 3, using the diffused UV–Vis reflection spectrum, the optical characteristics of the hBN, SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposites powder have been evaluated. With the exception of a small amount of absorption in the visible range (200–800 nm), SnO_2 mostly absorbs ultraviolet light. UV–visible spectra have been used to quantify a move in the bandgap energy of materials evenly


(a)
hBN
Tauc Plot
$(\alpha h\nu)^2$(eV cm^{-1})^2
E_g = 4.8 eV
Energy (eV)
Absorbance (a.u.)
Wavelength (nm)
(b)
SnO_2
Tauc Plot
$(\alpha h\nu)^2$(eV cm-1)^2
Eg = 3.65 eV
Energy (eV)
Absorbance (a.u.)
Wavelength (nm)
(c)
5% hBN-SnO_2
Tauc Plot
$(\alpha h\nu)^2$(eV cm^{-1})^2
E_g = 2.82 eV
Energy (eV)
Absorbance (a.u.)
Wavelength (nm)
(d)
7% hBN-SnO_2
Tauc Plot
$(\alpha h\nu)^2$(eV cm^{-1})^2
E_g = 2.72 eV
Energy (eV)
Absorbance (a.u.)
Wavelength (nm)
(e)
10% hBN-SnO_2
Tauc Plot
$(\alpha h\nu)^2$(eV cm^{-1})^2
Eg = 2.69 eV
Energy (eV)
Absorbance (a.u.)
Wavelength (nm)

◂**Figure 3** **a** UV–visible spectra of hBN flakes, **b** SnO_2 nanoparticle, **c** 5%hBN-SnO_2, **d** 7%hBN-SnO_2, and **e** 10%hBN-SnO_2 nanocomposites with optical band energy (Eg).

dispersed in ethanol–water. Boron nitride is a material having a high direct bandgap ($n = 2$) and the capacity to transition. Bulk hBN has the lowest value of E_g whereas h-BNNS holds the highest value of E_g [42]. A wider bandgap emphasizes quantum confinement in the substance and results in smaller crystallite sizes. The synthesized samples bandgap was also analysed using Tauc's figure;

$$(\alpha h\nu) = A(h - Eg)^{1/2} \quad (2)$$

In this equation, E_g is the bandgap, α is the absorption coefficient, A is a constant, and E is the photon energy obtained from $E = h\nu$, where h is Planck's constant, and ν is the UV–visible spectrum occurrence. Plotting $(\alpha h\nu)^2$ against $h\nu$ allowed for the calculation of the optical bandgap for pure hBN, pure SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposites. The calculated bandgap energies for pure SnO_2, hBN, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN were found to be 4.8 eV, 3.65 eV, 2.82 eV, 2.72 eV, and 2.69 eV. Once hBN is added to the SnO_2 nanoparticles, the bandgap value in the SnO_2-hBN nanocomposites drops, and the hump in the absorption spectrum keeps moving downward in wavelength.

Field emission scanning electron microscope

Figure 4 depicts the morphology of pure hBN, pure SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposite. It is feasible to verify that the morphology of the nanoflakes produced during the hydrothermal exfoliation of hBN was successfully synthesized using field emission scanning electron microscopy (FESEM). Figure 4a displaying FESEM image of hBN which is synthesized at reaction temperature 220 °C for 18 h displays large lateral size nanoflakes [43, 44]. Furthermore, aggregated hBN sheets are displayed at random. The FESEM picture of nanocrystalline SnO_2 is shown in Fig. 4b. It demonstrates the homogeneous and uniform dispersion of SnO_2 nanocrystallites [45]. Moreover, the FESEM image of the SnO_2-hBN (5%, 7%, and 10%) nanocomposite, as displayed in Fig. 4c–e, specifies that SnO_2 nanoparticles adhere consistently to the surface of hBN sheets. It indicates that hydrothermal treatment of hBN and $SnCl_4$ solution is a successful approach for the synthesis of the SnO_2-hBN nanocomposite [37]. The combination of SnO_2 with hBN nanocomposite demonstrates irregular shape particles combined with lateral size nanoflakes that could be favourable for electron transfer properties [43]. Energy-dispersive X-ray spectra, or EDX, are among the best methods available for determining sample purity. The SnO_2-hBN (5, 7, and 10%) nanocomposite is depicted in Fig. 5 and contains elements such as tin (Sn), oxygen (O), boron (B), and nitrogen (N), verified.

X-ray photoelectron spectroscopy

Furthermore, X-ray photoelectron spectroscopy (XPS) is employed to analyse the element configuration of the synthesized hBN, SnO_2, and SnO_2-hBN nanocomposite powder. In the absence of any other impurity, the survey spectra in Fig. 6a support the existence of B, N, C, Sn, and O. The high-resolution core level spectra of Sn 3d, B1s, N1s, C1s, and O1s are given in Fig. 6b–e.

B spectra

Figure 6b shows the B1s core level spectra for pure hBN and SnO_2-hBN nanocomposite, where high-resolution B1s spectrum with the main peak located at 190.2 eV which corresponds to BN bonds and a smaller peak at 190.6 eV assigned to BO bonds [43]. In the case of SnO_2-hBN nanocomposite powder, the majority of the B signal comes from the peaks at around 190 and 190.3 eV, which correspond to BN and BO, respectively [46]. The shift towards the lower binding energy side corresponds to electronic interaction at the interface of hBN and SnO_2 [47].

N spectra

Figure 6c displays the N1s core level spectra of the hBN and SnO_2-hBN nanocomposite. It is also possible to fit the broad peak surrounding the nitrogen 1 s lines with two distinct peaks at 397.9 and 398.2 eV, which are in good contact with the B–N and N–C bonds. These results show good agreement with the previous studies, suggesting that hBN nanosheets fabricated by the hydrothermal method display functional groups on their surface [48]. The deconvolution of N1s

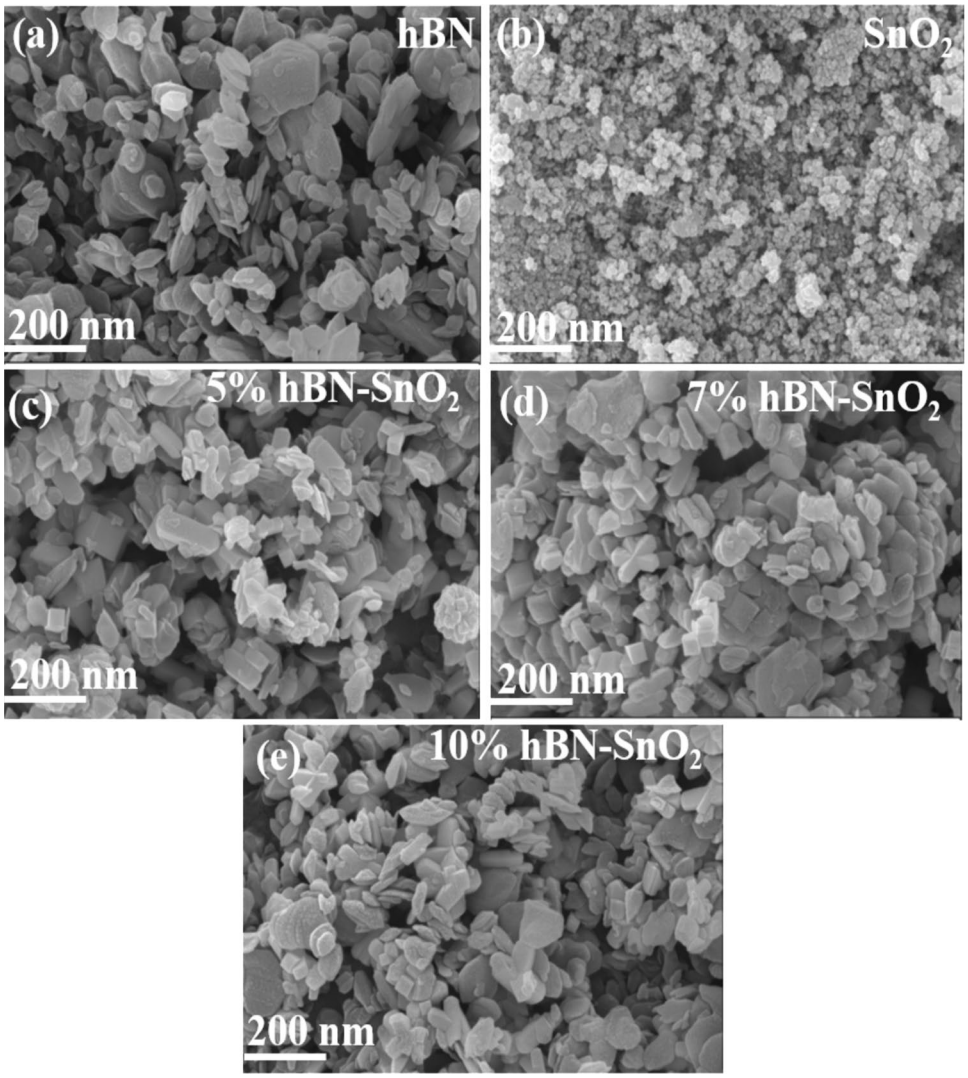


Figure 4 FESEM images of hBN, SnO_2, and hBN-SnO_2 nanocomposites with the different weight percentages of hBN: **a** hBN, **b** SnO_2, **c** 5%hBN-SnO_2, **d** 7%hBN-SnO_2, and **e** 10%hBN-SnO_2.

(SnO_2-hBN) spectra yielded two primary peaks at a binding energy of 397.6 and 397.8 eV (Fig. 6c). The peak located at 397.6 eV corresponds to the B–N bond of hBN whereas the peak located at lower energy (397.8 eV) can be attributed to the N–C bond, respectively [49].

Tin spectra

The Sn 3d positions of the SnO_2 and SnO_2-hBN compound samples are displayed in Fig. 6(d). In the case of pure SnO_2, two characteristic peaks of $Sn3d_{5/2}$ and $Sn3d_{3/2}$ are observed at 485.4 and 493.8 eV, respectively, which correspond to the typical oxidized form of Sn^{4+} [50, 51]. XPS peaks in the Sn 3d spectrum of the SnO_2-hBN nanocomposite show more substantial Sn $3d_{5/2}$ and Sn $3d_{3/2}$ peaks with binding energy changes of 1.3 eV, etc. These peaks are detected at 486.7 eV and 495.1 eV, respectively. The characteristic SnO_2 peaks are separated from one another by 8.4 eV, which is in line with the previous findings. The above-observed shift indicates that the SnO_2 nanoparticles are connected to the hBN nanosheets, and have also been confirmed in the literature [52]. A significant interfacial interaction between SnO_2 and hBN was required to improve resistive switching performance, as indicated by the formation of the Sn–O link between the two materials. As a result, the SnO_2-hBN and Sn 3d peaks have been moved to a higher binding energy position. Because SnO_2 has a smaller work function than hBN, the electrons in SnO_2-hBN hybrids can generally flow (SnO_2 to hBN) [20, 53]. Therefore, the hBN-SnO_2 Sn 3d peaks have been shifted to have a higher binding

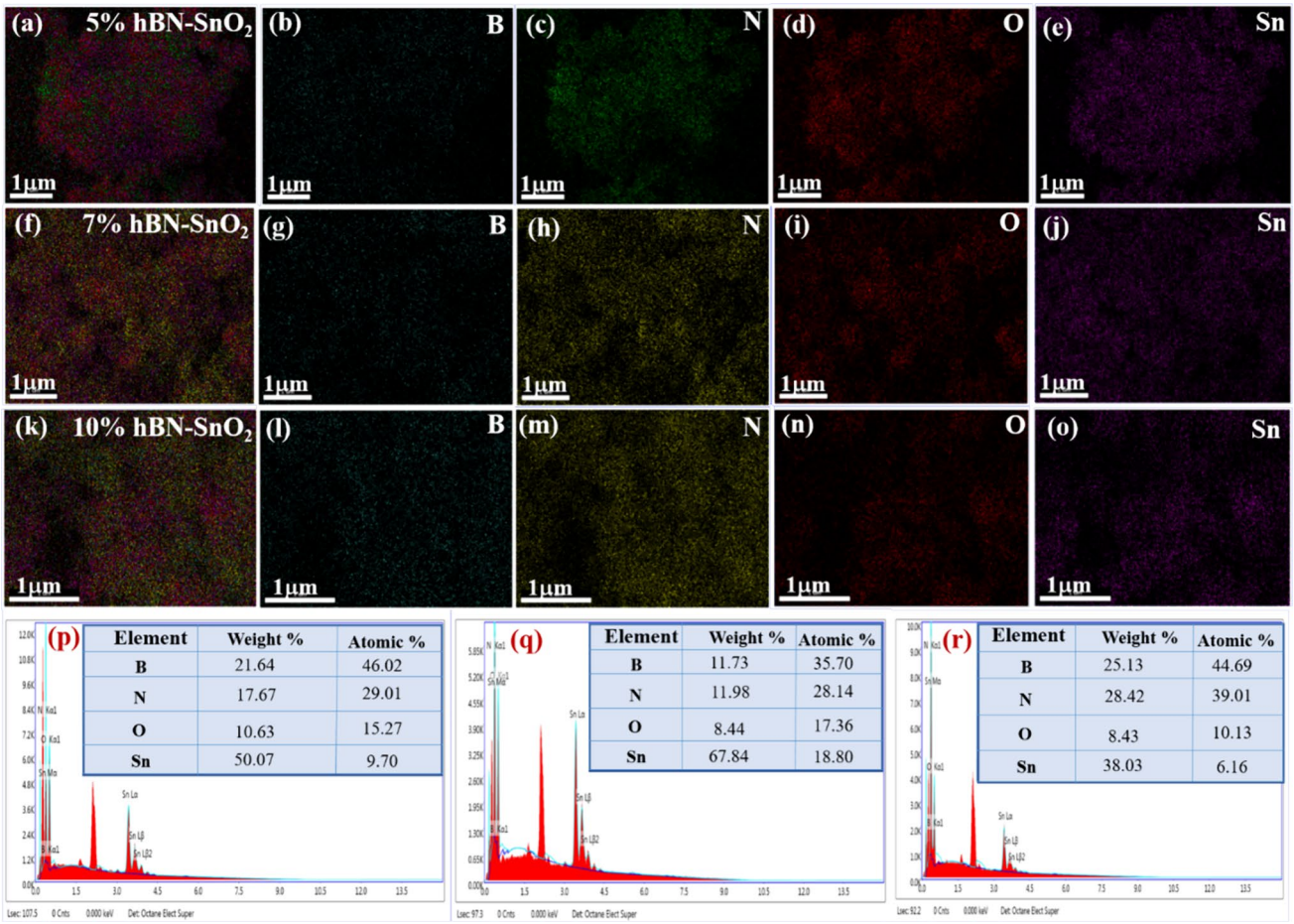


Figure 5 **a–o** Elemental mapping, **p–r** the synthesized EDX spectra of 5% hBN-SnO_2, 7% hBN-SnO_2, and 10% hBN-SnO_2 nanocomposites showing the purity of the synthesized powder.

energy. The electrostatic repulsive contact between O and a negative charge in the matrix allows Sn to generate more energy. The shift in the Sn electronic orbit will impact the binding energy of O1s [54]. Remarkably, when the SnO_2-hBN nanocomposite is synthesized, the associated peaks of B1s, N1s, and Sn 3d change positively, whilst the peaks of O1s move positively, indicating a significant interface among SnO_2 and hBN. The XPS results also show that ultrafine hBN flakes are agglomerated on the SnO_2 surfaces.

O1s spectra

The deconvoluted O1s core level spectra were found to consist of two peaks, as shown in Fig. 6e. Considering the information of lattice O, tetravalent Sn ions in the O vault are responsible for the peak for pure SnO_2 at the binding energy location of 530.3 eV. Another peak at 530.9 eV indicates that the surface is surrounded by hydroxide OH groups and corresponds to nonlattice oxygen [55]. Additionally, two peaks at 530.2 eV and 531.1 eV in the O1s spectra, corresponding to B–O and C–O bonds, are visible in the case of SnO_2-hBN, and they are displaced towards the lower binding energy side. BN from hBN and the oxygen species from SnO_2 likely interacted at the interface, as evidenced by the opposite direction shift in binding energy position for the O1s and BN peak [56].

Dielectric characteristics

Important details on the material relaxation method are revealed by the frequency dependence of the dielectric measurements. It gives details on the materials two primary physical properties, which are electric permittivity (sometimes referred to as the dielectric constant) and dielectric loss. A delay in polarity called

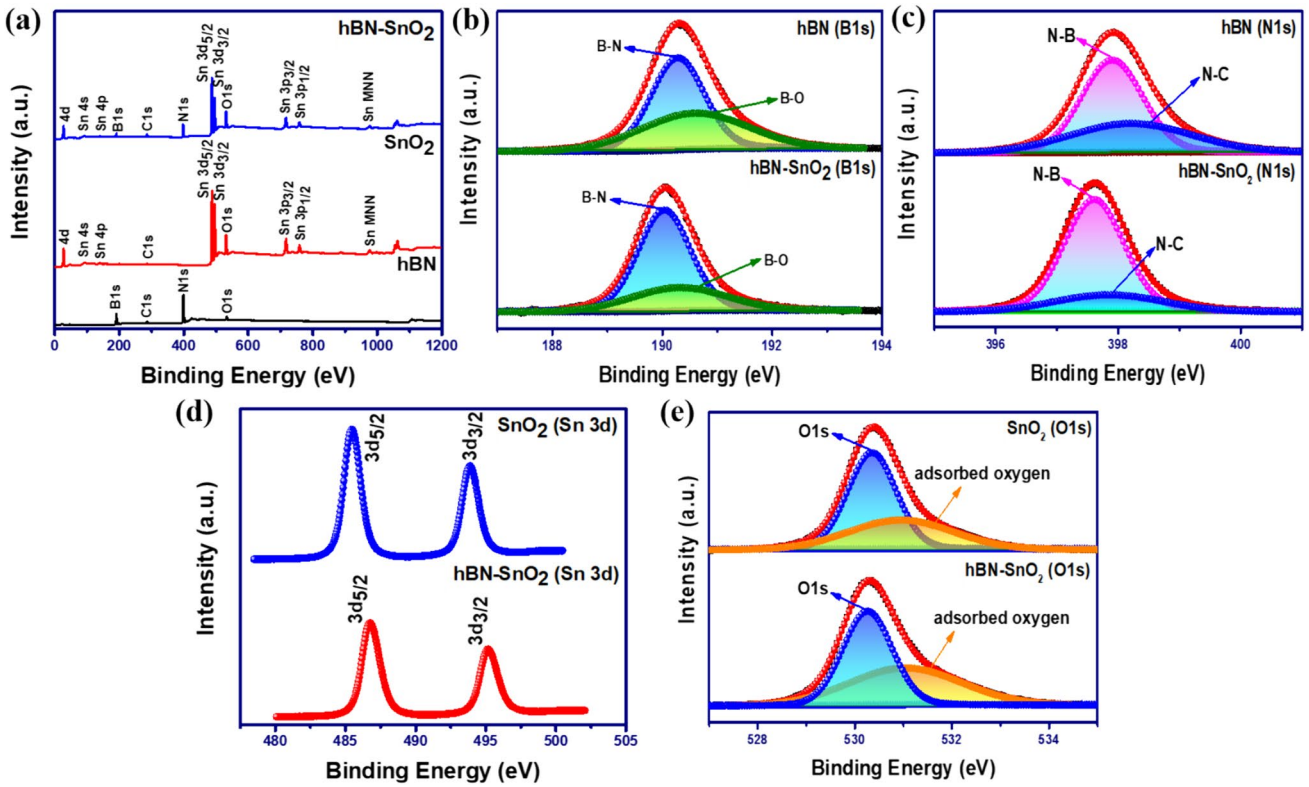


Figure 6 **a** An XPS survey scan of hBN-SnO_2 reveals the existence of elements such as B, N, Sn, and O, as well as increased resolution spectrums of **b** B1s, **c** N1s, **d** Sn 3d, and **e** O1s.

"dielectric loss" ($D = \tan\delta$) causes charge carriers to produce heat after a changing field is applied through the sample. The combined actions of electrical conduction and dielectric relaxation result in dielectric loss inside a material [57]. Figure 7a and b illustrates how the dielectric constant, also known as electric permittivity, varies with frequency in relation to its real ($\varepsilon' r$) and imaginary ($\varepsilon'' r$) components. Based on the Maxwell–Wagner model, it can be seen that $\varepsilon' r$ decreases with increasing frequency of applied field for all samples [58]. Space charge polarization enables nanodipoles that oscillate to facilitate easy orientation in the direction of the supplied electrical field; therefore, a high value of $\varepsilon' r$ is detected in the low-frequency zone [3]. Dipoles can travel to the grain boundaries at low frequencies, wherever charge carriers figure up and significantly polarize the material. The charge carriers are unable to keep up with the rapid changes in field direction that occurs at high frequencies. As a result, carriers are unable to reach the grain boundaries, and as a result, no polarizations occur [59]. Since composites have a larger surface area due to their small size (more grains), the dielectric constant in 7%SnO_2-hBN composite is higher than in 5%SnO_2-hBN, 10%SnO_2-hBN, pure SnO_2, and hBN [60]. In composites, more oxygen vacancies are formed, resulting in increased space charge polarization (SCP) and rotational direction polarization (RDP), which raises the dielectric constant. Furthermore, polarization occurs at the interface of SnO_2 and hBN, which contributes to the rise in the dielectric constant. This causes an increase in the dielectric constant value. Because hBN plays an important role in the alignment of the dipoles and facilitates the development of a microcapacitor within the composite. Furthermore, the hBN sheet has a very wide surface area, which causes a huge number of dipoles to form, improving charge generation and storage capability. As a result, the increase in dielectric constant seen following the addition of hBN could be attributed to the development of many micro-capacitors within the composite [61]. Furthermore, the presence of the interaction between SnO_2 and hBN, as well as the defects caused by the rearrangement of hBN sheets in the nanocomposite, generates free charge carriers in SnO_2-hBN, which accumulate at the interface and contribute to a strong interfacial polarization,

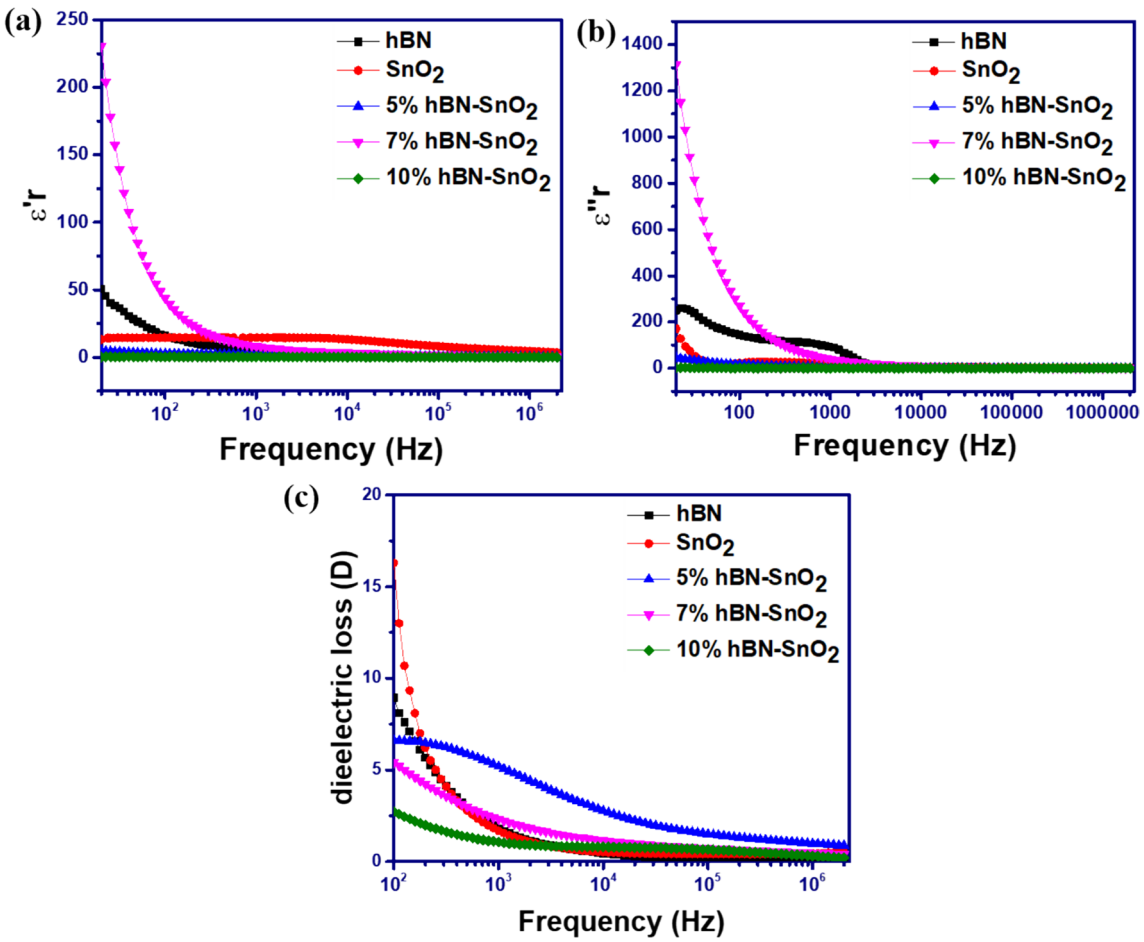


Figure 7 Dielectric constant variation **a** real and **b** imaginary parts, and dielectric loss (**c**) for pure hBN, SnO_2, 5% hBN-SnO_2, 7% hBN-SnO_2, and 10% hBN-SnO_2 nanocomposite as a function of the frequency.

resulting in an improvement in the composite's dielectric constant. It can also be shown from Fig. 7a and b that when hBN loading increases, the dielectric constant for SnO_2-hBN nanocomposite increases. However, once hBN loading reaches a threshold, the dielectric constant drops. Additionally, the highest dielectric constant is observed close to the percolation threshold, which is reached at the most efficient hBN adding in the SnO_2-hBN nanocomposite. In contrast, a tunnelling percolation method has been created within the material of the composite with higher hBN transferring beyond the percolation threshold, rendering the composites semiconductive [62]. Therefore, it is necessary to take into account charge injection at the contacts between metallic electrodes and composite materials. As a result, space charge polarization at the sample–electrode interfaces, together with interfacial polarization, reduces the dielectric constant.

A process called dielectric loss (tanδ) occurs in the dielectric system when energy is lost. Based on a frequency range of 20–10^6 Hz, Fig. 7c displays the dielectric loss for the room temperature nanocomposite of hBN, SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN. When the charge transporters in dielectric material fail to keep up with the changes in the electric field, dielectric loss occurs. For every sample, dielectric losses decrease as the applied field frequency increases. Due to domain wall inhibition-induced forced clockwise shift of magnetization, smaller electrical losses are observed at higher frequencies. At higher frequencies, lower values could indicate Maxwell–Wagner relaxation. As shown in Fig. 7c, lower frequencies cause the greatest dielectric loss. This is caused by space charge polarization, which can be explained by the Shockley–Read mechanism [63]. The variation of dielectric loss factor (tan δ) of pure hBN, 5%SnO_2-hBN, 7%SnO_2-hBN, and

10%SnO_2-hBN is discovered to be considerably lower than that of SnO_2. The dielectric constant (ε'') imaginary fraction denotes the quantity of energy dissipation, and tan δ is proportional to ε''. Figure 7c clearly shows that the SnO_2-hBN nanocomposite has low dielectric loss [58]. Furthermore, the decreased loss in the composite compared to SnO_2 could be attributable to significant interfacial polarization. In the case of SnO_2-hBN nanocomposite, the rise in dielectric loss may be due to the significant leakage current generated by the direct interaction between hBN sheets over the percolation threshold [59]. The high dielectric constant and low dielectric loss in the SnO_2-hBN sample may undoubtedly be attributed to the synergic impact of the hBN sheets and SnO_2 nanoparticle decoration in the composite. From Fig. 7c, it is observed that the dielectric loss tangent of SnO_2-hBN (5–10%) samples decreases from pure SnO_2 and pure hBN; thus, composites sample will be better in case of dielectric application [64, 65].

Impedance spectroscopy

To investigate numerous electrical properties, such as the impact of grain and grain borders on total impedance, relaxation time distribution, and the type of relaxation mechanism [57], the complicated behaviour of total impedance with frequency can be examined. Use the formula below to describe the whole composite impedance:

$$Z^* = Z' + jZ'' \tag{3}$$

wherever the real (Z′) and imaginary (Z′′) modules of the whole impedance are provided by Eq. 4.

$$Z'' = Z \times \sin\theta \quad \text{and} \quad Z' = Z \times \cos\theta \tag{4}$$

Here Z stands for the entire measured impedance or the observed phase angle (measured in radians). Figure 8a and b shows how the used field frequency

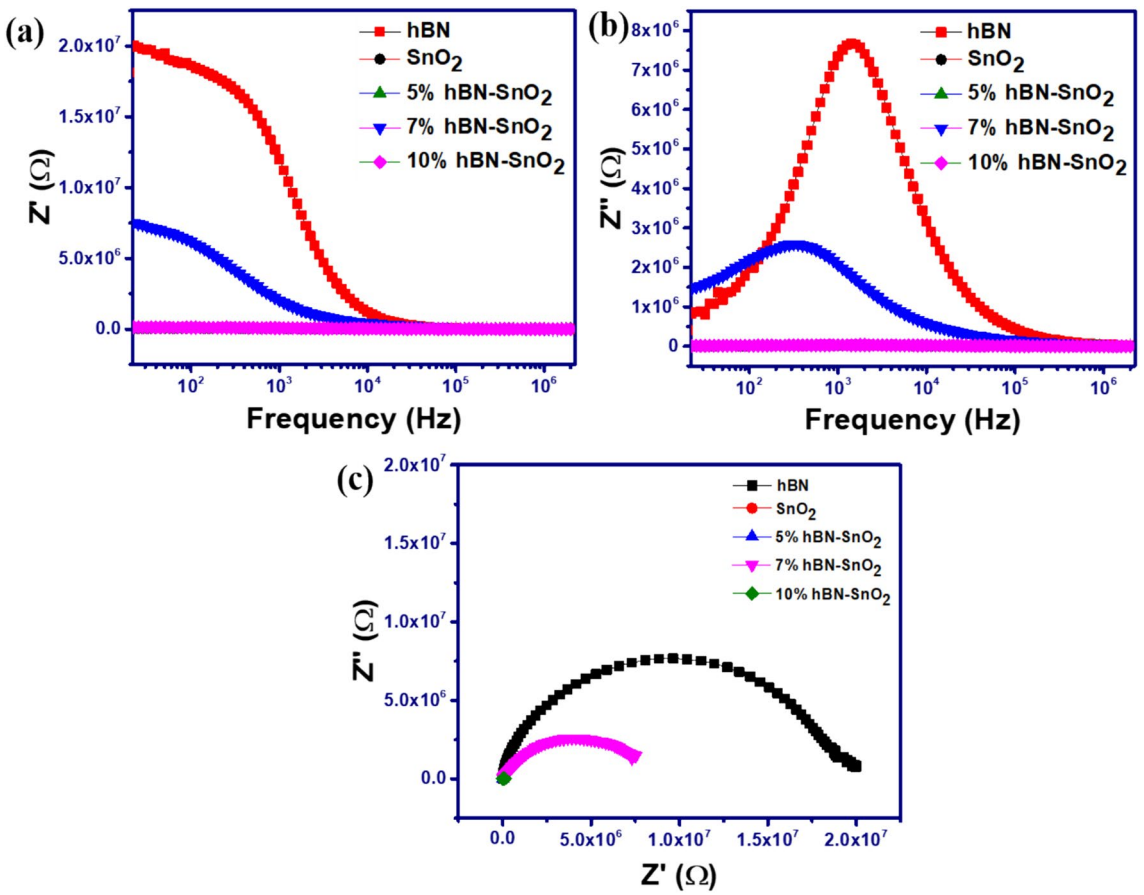


Figure 8 Impedance variation for the hBN, SnO_2, 5%hBN-SnO_2, 7%hBN-SnO_2, and 10%hBN-SnO_2 nanocomposite powder as a function of frequency (20–10^6 Hz) in both **a** real and **b** imaginary sections. **c** Nyquist plot of hBN, SnO_2, 5% hBN-SnO_2, 7% hBN-SnO_2, and 10% hBN-SnO_2 nanocomposite samples.

affects the fluctuation in the resistance or real part (Z') and the capacitance or imaginary part. Impedance levels at lower frequencies are considerable, but they get smaller as frequency rises. Because of increasing mobility and the dissipation of stationary charges, impedance has decreased as frequency has increased. At higher frequencies, there is a tendency for all the curves to merge due to a decrease in space charge and the cancellation of the dipole orientation effect [57]. Nyquist plots were used to separate the effects of grain boundaries at room temperature and grain temperature, as Fig. 8c illustrates. In materials used to create hBN, SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN nanocomposite, conductive grain boundaries are necessary at higher frequencies, whereas resistive grain boundaries are significant at lower frequencies. The semicircle shows exceptionally robust grain boundaries at low occurrences. Low resistance and high-frequency semicircles are produced by the conducting grains. In our situation, all of the samples display just one semicircle, which is caused by the contribution of the grain boundaries [64], as shown in Fig. 8c. Due to their higher resistance than the grain itself, grain boundaries are thought to play a significant role in impedance [66]. The SnO_2-hBN nanocomposite has less resistance than SnO_2 because of hBN strong conductivity.

AC conductivity

To better understand the conduction process, the frequencies-dependent AC electrical conductivity of hBN, SnO_2, 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN composites specimens was examined at ambient temperature in the occurrence range of 20–10^6 Hz. It has been noted that when the conductivity decreases with an increase in frequency, the free charge carriers make the majority contribution, whereas when the conductivity increases with frequency, the bound charge carriers make the majority contribution [67]. The carriers change from one faulty state to another as a result of the oscillating field over the sample. Figure 9a displays the calculated AC conductivity using the formula $\sigma(\omega) = \omega\varepsilon_0\varepsilon_0 \tan\delta$. It determines that conductivity rises with the occurrence, demonstrating the existence of bonded charge carriers and providing insight into the behaviour of considerable occurrence dependency. Carriers offer frequency-independent conductivity and can go farther since they can hop at lower frequencies for longer periods. Because carriers have shorter lifetimes, failure and successful processes conflict at higher frequencies. Higher frequency conductivity exhibits dispersion behaviour due to the ratio of failed to successful hops. Equation 5 is used to fit the frequency dependence of the AC conductivity data using Jonscher's power law [57].

$$\sigma_{ac}(\omega) = \sigma_{dc} + A\omega^s \tag{5}$$

where DC contribution is frequency-independent, s is a supporter describing the lattice–mobile ion interface based on the applied field occurrence, and A is the polarization strength. Moreover, $A\omega^s$ indicates the frequency-dependent contribution, whilst $\sigma_{ac}(\omega)$ reflects the overall observed conductivity. The shift between the frequency-independent and frequency-dependent regions demonstrates the transformation from long-range hopping across barriers to the short-range motion of carriers [57, 68]. An examination of AC conductivity reveals a sharp increase at higher frequencies, supporting Jonscher's power law and suggesting the potential of other conduction pathways. As shown in Fig. 9b–f, we estimate the AC conductivity values using Jonscher's power law. The fitting parameters derived from the data are presented in Table 2. Figure 9a shows that the addition of hBN to SnO_2 results in an increase in AC conductivity. Figure 9f shows that when the amount of hBN in SnO_2 increases, the AC conductivity decreases. Substituting hBN ions for SnO_2 ions hindered Verwey's hopping mechanism, resulting in reduced AC conductivity. Similar behaviour has already been observed by various researchers [69]. hBN ions at hexagonal boron and nitrogen sites serve a significant and dominant role in the conduction mechanism of 2D materials. SnO_2 with large ionic radii likes to be found on hexagonal sites. This reduces the quantity of hBN ions at the hexagonal boron and nitrogen sites, causing conductivity to drop [70].

I–V characteristics

The impact of hBN inclusion on SnO_2 electrical qualities is investigated using ambient temperature I–V characteristics measured with a Keithley 2450 Source Metre Unit. Figure 10 shows a diagram of the metal–insulator–metal (MIM) structure used to investigate I–V properties.

Figure 11a–d shows the I–V properties of Al/SnO_2/ITO-PET and Al/SnO_2-hBN/ITO-PET layers

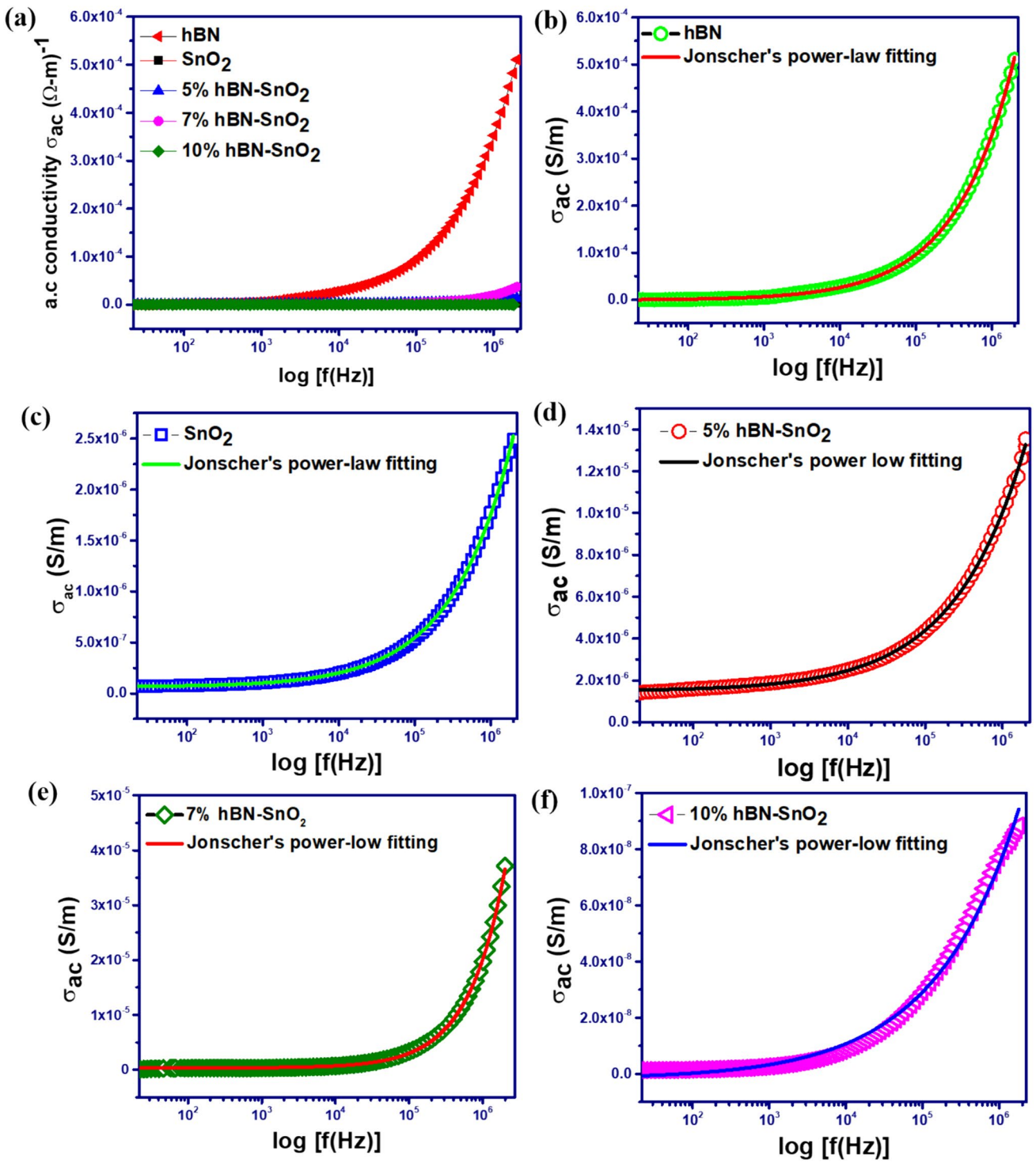


Figure 9 **a** The relationship between frequency and AC conductivity (σ_{ac}). **b–d** The graphs of curve fitting for hBN, SnO_2, 5% hBN-SnO_2, 7% hBN-SnO_2, and 10% hBN-SnO_2 nanocomposite were created using Jonscher's power law.

with different hBN wt.% in SnO_2 when the voltage is improved from − 1 to + 1 V. All of the fabricated memory devices resistive switching behaviour are shown in Fig. 11e–h. The device including the nanocomposite based on SnO_2-hBN exhibits distinct behaviour from the pure SnO_2-based device,

Table 2 Parameters are derived from Jonscher's power law fitting for each sample

Materials	Frequency exponents (s)	Pre-exponent A
SnO_2	0.543	9.22×10^{-10}
hBN	0.558	1.56×10^{-7}
5% hBN-SnO_2	0.467	1.33×10^{-8}
7% hBN-SnO_2	0.870	1.17×10^{-10}
10% hBN-SnO_2	0.391	3.40×10^{-10}

as demonstrated, suggesting the presence of interaction at the SnO_2-hBN interface. To investigate the memory phenomenon, the power is raised. It is found that the device enters the SET state (6.5 V) for pure SnO_2-based devices (Al/SnO_2/ITO-PET), transitioning from the early HRS to the LRS. However, whilst the voltage continues to rise in the same direction, as shown in Fig. 11e, the changeover from LRS to HRS does not occur. For the device to complete the reset transition, it is necessary to reverse the voltage polarity. The bipolar resistive switching behaviour is demonstrated by the device, which has been found to occur at − 5.5 V. The direction of the voltage sweep is shown by the arrow. SnO_2-based memory devices are not suited for real-world applications because of their greater switching voltages and less hysteresis. Consequently, more research was done on the device using an active layer made of SnO_2-hBN nanocomposite. To further comprehend the influence of 2D hBN in SnO_2 on memory device applications, various wt.% of hBN in SnO_2-based devices were investigated. Every SnO_2-hBN nanocomposite device exhibited the properties of bipolar resistive switching (BRS), as shown in Fig. 11f–h. Moreover, as shown in Fig. 11g, the memory phenomena with 7 wt.% hBN in SnO_2 show a greater ON/OFF current ratio. When comparing the nanocomposite-based memory device to a pure SnO_2-based memory device, the HRS to LRS transition occurs at a voltage of 0.5 V, whereas the LRS to HRS transition occurs at -0.7 V. The device made of SnO_2-hBN has better resistive switching characteristics. The I_{ON}/I_{OFF} ratios at a read voltage of 0.5 V are 2, 96, 1000, and 10 for pure SnO_2, 5%, 7%, and 10% SnO_2-hBN devices, respectively. This suggests that the composite 7 wt.% hBN-SnO_2 device performs better when it comes to resistive switching. After comparing to pure SnO_2, the interfacial contact between hBN and SnO_2 nanoparticles may explain the greater resistive switching ratio (I_{ON}/I_{OFF}). The dielectric constant of a substance determines its ability to hold charges and has been found that a device having a 7%SnO_2-hBN nanocomposite has a higher dielectric constant, which allows for additional charge storage, potentially extending the memory window. It should be mentioned that the composite made of SnO_2 and hBN

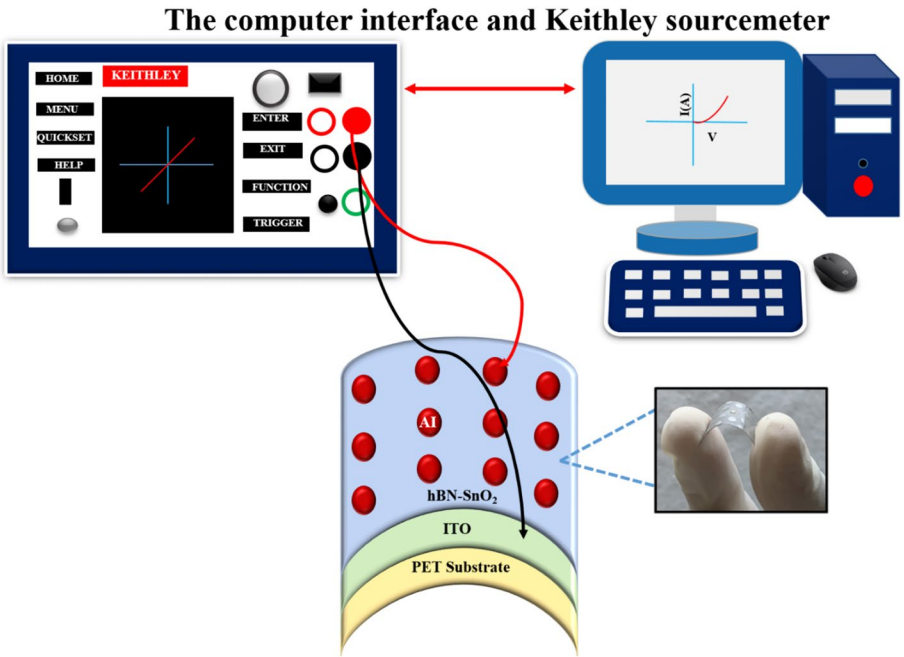
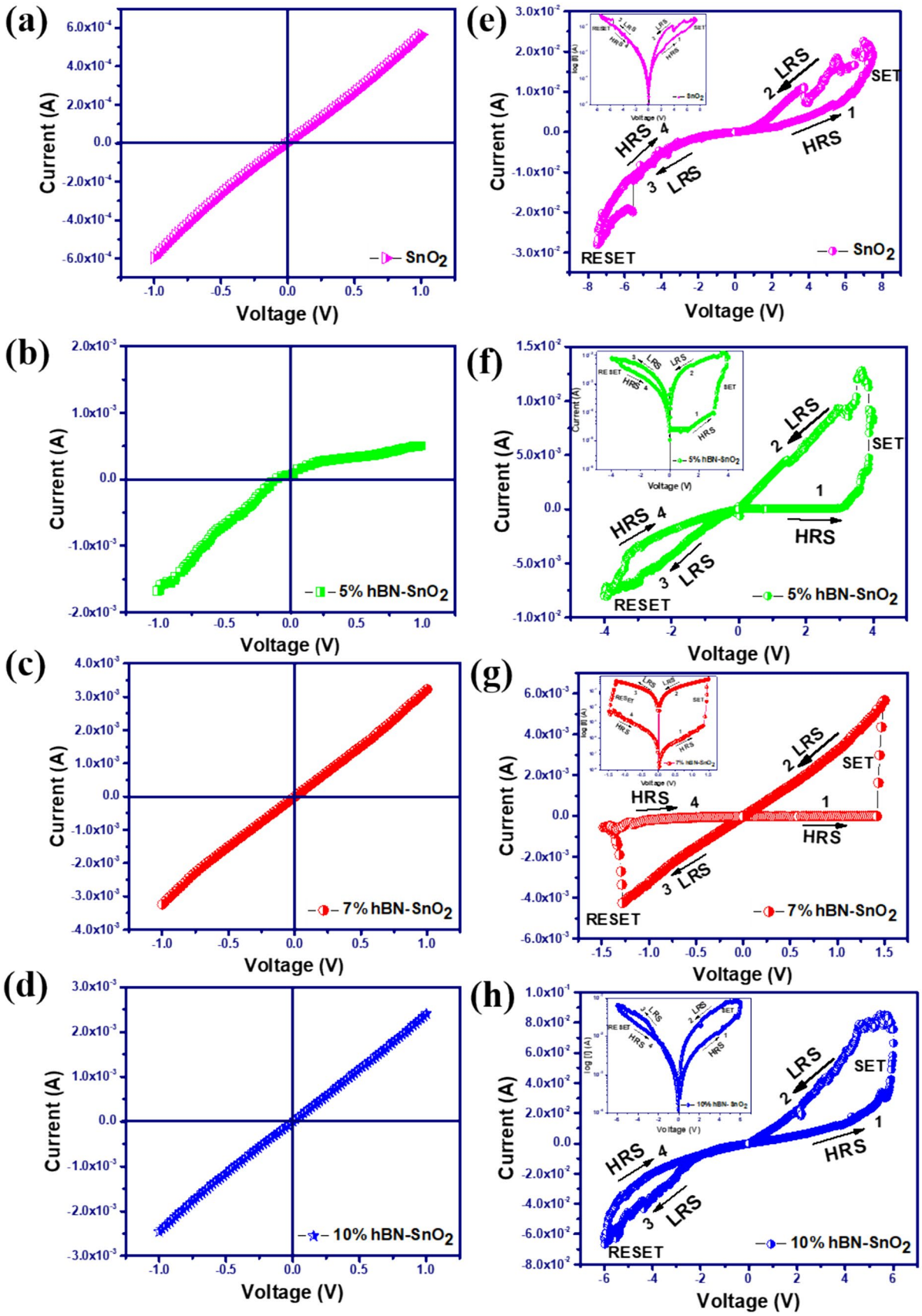


Figure 10 Schematic depicting the measuring procedure used an optical image of the actual memory device and a Keithley 2450 Source Metre with a LabVIEW-controlled programme for I–V measurement.


(a)
Current (A)
Voltage (V)
SnO2
(b)
Current (A)
Voltage (V)
5% hBN-SnO2
(c)
Current (A)
Voltage (V)
7% hBN-SnO2
(d)
Current (A)
Voltage (V)
10% hBN-SnO2
(e)
Current (A)
Voltage (V)
LRS
SET
HRS
RESET
SnO2
(f)
Current (A)
Voltage (V)
LRS
SET
HRS
RESET
5% hBN-SnO2
(g)
Current (A)
Voltage (V)
LRS
SET
HRS
RESET
7% hBN-SnO2
(h)
Current (A)
Voltage (V)
LRS
SET
HRS
RESET
10% hBN-SnO2

◂**Figure 11** **a–d** The I–V properties and **e–h** the resistive switching properties observed for pure SnO_2 and hBN-SnO_2 nanocomposite with the different weight percentages of hBN powder. The inset of resistive switching figures shows the logarithmic I–V curves.

also considerably increases the switching parameter, as seen by the XPS spectra. According to Singh et al., resistive memory-based metal oxide nanocomposite applications have two advantages: They not only serve as an oxygen reservoir but also as a blocking layer to stop oxygen from diffusing into the switching layers, improving the switching ratio [71].

As shown in Fig. 12a, the positive side characteristics of I–V are presented on a dual logarithmic scale to help better understand the switching operations in Al/7%SnO_2-hBN/ITO-PET. Because the device-fitted I–V curve in the HRS has a slope of 1, the ohmic conduction mechanism dominates at lower voltages, as demonstrated in Fig. 12a. The density of thermally generated charged particles in dielectric materials is significantly higher than the density of introduced charge particles since metal electrodes at lesser voltage values, as per the ohmic conduction process. A dielectric or insulator's electrical properties and the applied voltage influence how much generated or inserted charge is produced by metallic electrodes. The charge transporters produced by metal electrodes frequently pass over the entire depth of the sample due to the long relaxation period. This state is referred to as high resistance (HRS). Injecting additional electricity from the metal electrode results in a rise in the applied voltage. The quantity of charge carriers injected exceeds the quantity of charge carriers produced by heat. All available trap sites are filled with induced or injected charge carriers, which causes a large rise in current. In this area, the present correlates to the square of the voltage, and space charge-limited conduction (SCLC) has replaced ohmic conduction. The induced charge carriers move through the materials upon filling every vacant trap site. Current flowing through the material increases dramatically as a result [69] (see Fig. 12a). Subsequently, due to space charge-limited conduction, injected electrons fill every trap in the active film of the 7%SnO_2-hBN material, and the current is determined by squaring the voltage. Thus, in the high-voltage area, the slope rises close to 2.5 based on the I–V curves. This is confirmed by applying the Schottky emission concept to the high-voltage area I–V characteristics (a slope of 2.5 on the log (I) vs. log (V) graph). The resulting SnO_2-hBN nanocomposites and the electrode differing work functions, as indicated by the equation below, are what cause this interface-limited conduction mechanism.

$$\ln I \propto \sqrt{\frac{e^3}{4\pi\varepsilon_r\varepsilon_0 d}}\sqrt{V}/kT \tag{6}$$

In the higher voltage area of HRS, Schottky emission is the predominant charge transport method, as indicated by the linear plot between log (I) versus $V^{1/2}$ displayed in the inset of Fig. 12a. The excellent fit of the I–V curve at a higher voltage by the Schottky model indicates that multiple charge transport mechanisms are operational in HRS [1, 72]. Additionally, the entire region of LRS is equipped with an ohmic conduction slope of roughly 1.2. This type of mechanism is seen in devices where the contact at the junction is ohmic, allowing carriers to enter and move freely inside the insulating layer [73]. When a material conduction band is injected with electrons more quickly than they are reabsorbed, the SCLC mechanism takes place. This huge amount of electrons limits the devices current flow due to electron scattering. The HRS area, which comprises this SCLC mechanism, may be caused by oxygen vacancies within the hBN thin layer.

A sequence of durability examinations was carried out to assess the device constancy with the largest memory window (Al/7%SnO_2-hBN/ITO-PET). Figure 12b shows that the device performance remains stable even after 100 cycles. The set and reset functions were performed with sweeping voltages ranging from 0 to + 1.5 V and 0 to − 1.5 V, respectively. At ambient temperature, we used a 0.5-V read voltage to evaluate the resistance across ON and OFF. The retention parameters displayed in Fig. 12d demonstrate that HRS and LRS may be retained by the 7%SnO_2-hBN-based device for more than 1.5×10^4 s without suffering any significant degradation. As illustrated in Fig. 12c, the device memory window is 1000, as represented by R_{ON}/R_{OFF}. It is obvious from these results that constant after several cycles, the Al/7%SnO_2-hBN/ITO-PET device does not degrade. According to these findings, the change from the OFF to the ON state is reversible and repeatable. At the read voltage of 0.5 V, both HRS and LRS remain unchanged when the resistance ratio of LRS/HRS stays constant throughout 100 cycles. According to the results overhead,

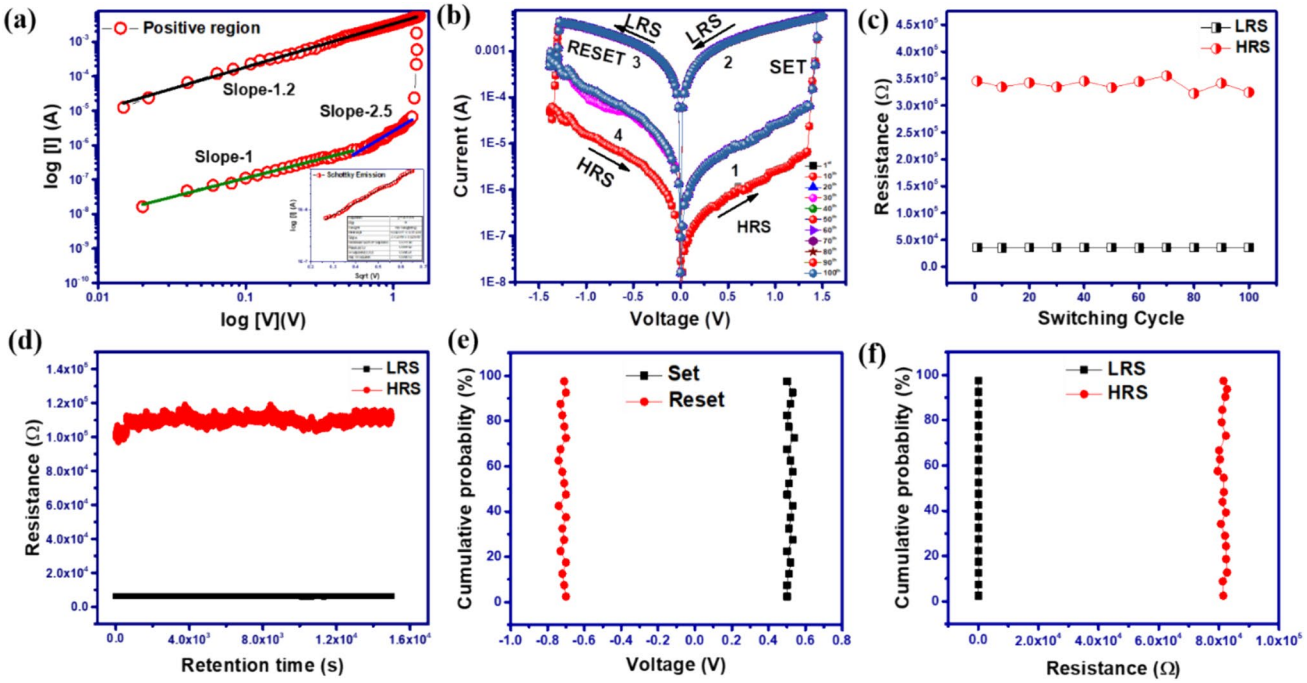


Figure 12 **a** Positive LRS, HRS at log–log scale. **b**, **c** The hBN-SnO_2-based device endurance performance over 100 resistive switching cycles. **d** Sustained resistance in HRS and LRS for roughly 1.6×10^4 s stipulating robust performance. **e**, **f** The device-to-device cumulative probability distributions of Vset–Vreset and HRS-LRS of 7%hBN-SnO_2 RRAM devices.

the Al/7%SnO_2-hBN/ITO-PET device performs consistently and durably, and switching between HRS and LRS is improved. We also examined device-to-device continuous probability distributions for the $V_{SET} - V_{RESET}$, LRS, and HRS for the 7%SnO_2-hBN RRAM devices, which are displayed in Fig. 12e, f, and so on. The operating voltages for the device-to-device distribution were calculated using the average of the first 10 cycles of each of the 20 tested devices [74]. Since the filament rupture process can be unpredictable, it is vital to handle the little variance in the RESET process for a practical memory application. Across devices, there is virtually little variation in the SET voltage variability. The fact that the device-to-device LRS and HRS vary very little should be addressed as they demonstrate the fabricated memory devices dependability [75].

Summary

In conclusion, an Al/SnO_2-hBN/ITO-PET-based memory device has been fabricated in the current study, where the percentage of hBN was varied to understand its influence on resistive switching properties. The nanocomposite powder with 5%SnO_2-hBN, 7%SnO_2-hBN, and 10%SnO_2-hBN with average crystallize sizes of 26.1 nm, 20.8 nm, and 20.7 nm is successfully synthesized via hydrothermal synthesis. XRD, Raman, FTIR, FESEM, EDX, and XPS analysis are used systematically to investigate the phase structure, composite formation, surface structure, chemical distribution, and valance state of SnO_2-hBN nanocomposite. When comparing the SnO_2-hBN nanocomposite to pure SnO_2 and hBN, the shift in the B1s, N1s, C1s, O1s, and Sn 3d peaks indicates an interfacial interaction. The dielectric constant, dielectric loss factor, and AC conductivity were found to decrease with increasing hBN quantity and frequency. In comparison with SnO_2, 5%SnO_2-hBN, and 10%SnO_2-hBN nanocomposites, the 7%SnO_2-hBN nanocomposite had a higher dielectric constant and lower loss. The percolation theory of conductive filler may explain dramatic variations in the dielectric characteristics of SnO_2-hBN. One semicircle is displayed in the Nyquist plots only because of the grain boundary contribution to the conduction process. AC conductivity was shown to be substantially dependent on frequency, with conductivity increasing at higher frequencies. The power law of Jonscher's is universally applicable to AC

conductivity. Later, the fabricated devices made up of pure SnO_2 and SnO_2-hBN (5, 7, and 10 wt.%) nanocomposite powder were tested for resistive switching applications. The resistive memory device based on SnO_2-hBN nanocomposite further demonstrated its altered electrical properties with a decrease in switching parameters and an increase in the current ratio. After comparing the resistive switching performance of the created device, it was found that the device with 7% hBN in SnO_2 had a higher I_{ON}/I_{OFF} ratio of ~ 1000, whereas the devices with 5 and 10 wt.% hBN in SnO_2 had higher ratios of ~ 96 and ~ 10, respectively. The Al/7%SnO_2-hBN/ITO-PET device also showed good durability behaviour up to 100 cycles and switching characteristics with a lower switching voltage. Retention time, which is an important consideration for RRAM devices, was measured for this device up to 1.5×10^4 s. The increased electrical properties of the presented flexible device indicate possibilities for future RRAM memory.

Acknowledgements

Thank you to the Government of India CSIR for sponsoring the scholarship and to AIRF (JNU) for helping with the Raman spectroscopy measurement.

Author contributions

KK helped in methodology, writing—original draught, investigation, conceptualization, data curation. BS contributed to writing—review & editing, supervision, methodology, conceptualization. MS contributed to revising the manuscript critically to enhance its crucial intellectual content and visualization.

Data and code availability

The data presented in this article will be available on reasonable request.

Declarations

Conflict of interest A conflict of interest is not present, according to the authors.

References

[1] Zhu S, Sun B, Zhou G, Guo T, Ke C, Chen Y, Yang F, Zhang Y, Shao J, Zhao Y (2023) In-depth physical mechanism analysis and wearable applications of HfO x-based flexible memristors. ACS Appl Mater Interfaces 15:5420–5431

[2] Wang D, Zhao S, Li L, Wang L, Cui S, Wang S, Lou Z, Shen G (2022) All-flexible artificial reflex arc based on threshold-switching memristor. Adv Funct Mater 32:2200241

[3] Hou ZL, Ma X, Zhang J, Li C, Wang Y, Cao M (2022) Fascinating electrical transport behavior of topological insulator Bi_2Te_3 nanorods: toward electrically responsive smart materials. Small 18:2205624

[4] Gergel-Hackett N, Hamadani B, Dunlap B, Suehle J, Richter C, Hacker C, Gundlach D (2009) A flexible solution-processed memristor. IEEE Electron Device Lett 30:706–708

[5] Lian H, Cheng X, Hao H, Han J, Lau M-T, Li Z, Zhou Z, Dong Q, Wong W-Y (2022) Metal-containing organic compounds for memory and data storage applications. Chem Soc Rev 51:1926–1982

[6] Chen Y, Yan Y, Wu J, Wang C, Lin JY, Zhao JS, Hwang CS (2020) Electroforming-free, flexible, and reliable resistive random-access memory based on an ultrathin TaOx film. ACS Appl Mater Interfaces 12:10681–10688

[7] Huang Y, Gu Y, Mohan S, Dolocan A, Ignacio ND, Kutagulla S, Matthews K, Londoño-Calderon A, Chang YF, Chen YC (2023) Reliability improvement and effective switching layer model of thin-film MoS_2 memristors. Adv Funct Mater 34:2214250

[8] Ahmadi M, Zabihi O, Jeon S, Yoonessi M, Dasari A, Ramakrishna S, Naebe M (2020) 2D transition metal dichalcogenide nanomaterials: advances, opportunities, and challenges in multi-functional polymer nanocomposites. J Mater Chem A 8:845–883

[9] Feng P, Chao C, Wang Z-S, Yang Y-C, Jing Y, Fei Z (2010) Nonvolatile resistive switching memories-characteristics, mechanisms and challenges. Prog Nat Sci Mater Int 20:1–15

[10] Jeong H, Kim J, Kim DY, Kim J, Moon S, Ngome Okello OF, Lee S, Hwang H, Choi S-Y, Kim JK (2020) Resistive switching in few-layer hexagonal boron nitride mediated by defects and interfacial charge transfer. ACS Appl Mater Interfaces 12:46288–46295

[11] Varun I, Mahato AK, Raghuwanshi V, Tiwari SP (2020) Ultralow current switching in flexible hybrid PVP: MoS_2/HfO x bilayer devices. IEEE Trans Electron Devices 67:3472–3477

[12] Nagareddy VK, Barnes MD, Zipoli F, Lai KT, Alexeev AM, Craciun MF, Wright CD (2017) Multilevel ultrafast flexible nanoscale nonvolatile hybrid graphene oxide–titanium oxide memories. ACS Nano 11:3010–3021

[13] Sun Y, Wen D (2019) Nonvolatile WORM and rewritable multifunctional resistive switching memory devices from poly (4-vinyl phenol) and 2-amino-5-methyl-1, 3, 4-thiadiazole composite. J Alloy Compd 806:215–226

[14] Bhattacharjee S, Das U, Sarkar PK, Roy A (2018) Stable charge retention in graphene-MoS_2 assemblies for resistive switching effect in ultra-thin super-flexible organic memory devices. Org Electron 58:145–152

[15] Wang H, Wang N, Cheng T, Wan X, Sun R, Wong CP (2019) Halides-assisted low-temperature synthesis of hexagonal boron nitride nanosheets. Part Part Syst Charact 36:1900278

[16] Chen J, Huang X, Sun B, Jiang P (2018) Highly thermally conductive yet electrically insulating polymer/boron nitride nanosheets nanocomposite films for improved thermal management capability. ACS Nano 13:337–345

[17] Siddiqui GU, Rehman MM, Yang Y-J, Choi KH (2017) A two-dimensional hexagonal boron nitride/polymer nanocomposite for flexible resistive switching devices. J Mater Chem C 5:862–871

[18] Yuan R, Xia W, Xu M, Miao Z, Wu S, Zhang X, He J, Wang Q (2020) Rectifying resistance switching behaviors of SnO_2 microsphere films modulated by top electrodes. Curr Appl Phys 20:431–437

[19] Boratto MH, Ramos RA Jr, Congiu M, Graeff CF, Scalvi LV (2017) Memristive behavior of the SnO_2/TiO_2 interface deposited by sol–gel. Appl Surf Sci 410:278–281

[20] Komal K, Gupta G, Singh M, Singh B (2022) Improved resistive switching of RGO and SnO_2 based resistive memory device for non-volatile memory application. J Alloy Compd 923:166196

[21] Nieh C, Lu M, Weng T, Chen Y (2014) Resistive memory of single SnO_2 nanowire based switchable diodes. Appl Phys Lett 104:213501. https://doi.org/10.1063/1.4880210

[22] Singh S, Sattigeri RM, Kumar S, Jha PK, Sharma S (2021) Superior room-temperature ammonia sensing using a hydrothermally synthesized MoS_2/SnO_2 composite. ACS Omega 6:11602–11613

[23] Trivedi S, Joshi U (2017) Resistive switching properties of highly transparent SnO [2]: Fe. Nano- Electron. Phys. 9(1):01025. https://doi.org/10.21272/jnep.9(1).01025

[24] Komal K, Singh M, Singh B (2023) One step hydrothermal synthesis of MoS_2–SnO_2 nanocomposite for resistive switching memory application. J Mater Sci Mater Electron 34:1351

[25] Komal K, Singh M, Singh B (2023) Flexible SnO_2–MoS_2 based memristive device exhibiting stable and enhanced memory phenomenon. J Phys D Appl Phys 57:105107

[26] Jung J, Shin D, Lee Y, Pak JJ (2021) Fabrication of solution-processed SnO_2–based flexible ReRAM using laser-induced graphene transferred onto PDMS. Curr Appl Phys 25:70–74

[27] Yun MJ, Kim KH, Bea D, Jung J, Kim S, Kim H-D (2021) Improved resistive switching of SnO_2 based resistive random access memory devices using post microwave treatment. J Electr Eng Technol 16:1011–1017

[28] Sabzevar PS, Bagheri-Mohagheghi M, Shirpay A (2023) SnO2: rGO transparent semiconducting thin films under annealing by hydrazine—modification of optical gap and electrical resistance. J Mater Sci: Mater Electron 34:791

[29] Zhang D, Sun Y, Li P, Zhang Y (2016) Facile fabrication of MoS_2-modified SnO_2 hybrid nanocomposite for ultrasensitive humidity sensing. ACS Appl Mater Interfaces 8:14142–14149

[30] Ni X, Chen C, Wang Q, Li Z (2019) One-step hydrothermal synthesis of SnO_2-MoS_2 composite heterostructure for improved visible light photocatalytic performance. Chem Phys 525:110398

[31] Sun B, Zhao W, Liu Y, Chen P (2015) Resistive switching effect of Ag/MoS_2/FTO device. Funct Mater Lett 8:1550010

[32] Le Trung H, Tran NK, Manh HC, Duc HN, Quynh HTT, Pham TH (2018) Hydrothermal synthesis, structure, and photocatalytic properties of SnO_2/rGO nanocomposites with different GO concentrations. Mater Res Express 5:095506

[33] Zhang D, Chang H, Liu R (2016) Humidity-sensing properties of one-step hydrothermally synthesized tin dioxide-decorated graphene nanocomposite on polyimide substrate. J Electron Mater 45:4275–4281

[34] Nath D, Mandal S, Deb D, Rakshit J, Dey P, Roy J (2018) Light tuning DC and AC electrical properties of ZnO-rGO based hybrid nanocomposite film. J Appl Phys 123:095115

[35] Jain N, Sharma S, Puri NK (2021) Investigation of charge transport mechanism in hydrothermally synthesized reduced graphene oxide (rGO) incorporated zinc oxide (ZnO) nanocomposite films. J Mater Sci Mater Electron 33:1307–1323

[36] Sahu BK, Pal A, TS R, Das A (2022) Single-molecule detection-enabled plasmonic Ag nanogap for unmasking vibrational properties in 0D SnO_2. ACS Appl Nano Mater 5:12413–12422

[37] Komal K, Singh M, Singh B (2023) Flexible SnO_2-MoS_2 based memristive device exhibiting stable and enhanced memory phenomenon. J Phys D Appl Phys 57:105107

[38] Saleh S, Ibrahim A, Mohamed S (2016) Structural and optical properties of nanostructured Fe-doped SnO_2. Acta Phys Pol A 129:1220–1225

[39] Zhang H-X, Shin B-G, Lee D-E, Yoon K-B (2020) Preparation of PP/2D-nanosheet composites using $MoS_2/MgCl_2$- and $BN/MgCl_2$-bisupported ziegler-natta catalysts. Catalysts 10:596

[40] Ayeshamariam A, Ramalingam S, Bououdina M, Jayachandran M (2014) Preparation and characterizations of SnO_2 nanopowder and spectroscopic (FT-IR, FT-Raman, UV–Visible and NMR) analysis using HF and DFT calculations. Spectrochim Acta Part A Mol Biomol Spectrosc 118:1135–1143

[41] Shenoy MR, Ayyasamy S, Bhojan V, Swaminathan R, Raju N, Senthil Kumar P, Sasikumar M, Kadarkarai G, Tamilarasan S, Thangavelu S (2021) Visible light sensitive hexagonal boron nitride (hBN) decorated Fe_2O_3 photocatalyst for the degradation of methylene blue. J Mater Sci Mater Electron 32:4766–4783

[42] Sharma K, Puri NK (2021) Enhanced electrochemical performance of hydrothermally exfoliated hexagonal boron nitride nanosheets for applications in electrochemistry. J Electrochem Soc 168:056512

[43] Kesavan G, Gopi PK, Chen S-M, Vinothkumar V (2021) Iron vanadate nanoparticles supported on boron nitride nanocomposite: electrochemical detection of antipsychotic drug chlorpromazine. J Electroanal Chem 882:114982

[44] An L, Bai C, Bai Y, Zhang B, Yu Y, Hu Y (2020) Superhydrophobicity of polyvinylidene fluoride induced by integrating liquid-exfoliated hexagonal boron nitride nanosheets. High Perform Polym 32:73–82

[45] Sharma A, Arya S, Chauhan D, Solanki PR, Khajuria S, Khosla A (2020) Synthesis of Au–SnO_2 nanoparticles for electrochemical determination of vitamin B12. J Mark Res 9:14321–14337

[46] Jerome R, Sundramoorthy AK (2020) Preparation of hexagonal boron nitride doped graphene film modified sensor for selective electrochemical detection of nicotine in tobacco sample. Anal Chim Acta 1132:110–120

[47] Bamola P, Singh B, Bhoumik A, Sharma M, Dwivedi C, Singh M, Dalapati GK, Sharma H (2020) Mixed-phase TiO_2 nanotube–nanorod hybrid arrays for memory-based resistive switching devices. ACS Applied Nano Mater 3:10591–10604

[48] Chen Y, Xu X, Li C, Bendavid A, Westerhausen MT, Bradac C, Toth M, Aharonovich I, Tran TT (2021) Bottom-up synthesis of hexagonal boron nitride nanoparticles with intensity-stabilized quantum emitters. Small 17:2008062

[49] Rastogi PK, Sahoo KR, Thakur P, Sharma R, Bawari S, Podila R, Narayanan TN (2019) Graphene–hBN non-van der Waals vertical heterostructures for four-electron oxygen reduction reaction. Phys Chem Chem Phys 21:3942–3953

[50] Jadhav H, Suryawanshi S, More M, Sinha S (2017) Pulsed laser deposition of tin oxide thin films for field emission studies. Appl Surf Sci 419:764–769

[51] Liang Y, Fang B (2013) Hydrothermal synthesis of SnO_2 nanorods: morphology dependence, growth mechanism and surface properties. Mater Res Bull 48:4118–4124

[52] Li S, Xie W, Wang S, Jiang X, Peng S, He D (2014) Facile synthesis of rGO/SnO_2 composite anodes for lithium ion batteries. J Mater Chem A 2:17139–17145

[53] Yang D, Yang R, Wang K, Wu C, Zhu X, Feng J, Ren X, Fang G, Priya S, Liu SF (2018) High efficiency planar-type perovskite solar cells with negligible hysteresis using EDTA-complexed SnO_2. Nat Commun 9:1–11

[54] Wang Z, Han T, Fei T, Liu S, Zhang T (2018) Investigation of microstructure effect on NO_2 sensors based on SnO_2 nanoparticles/reduced graphene oxide hybrids. ACS Appl Mater Interfaces 10:41773–41783

[55] Babu B, Reddy IN, Yoo K, Kim D, Shim J (2018) Bandgap tuning and XPS study of $SnO_2$2 quantum dots. Mater Lett 221:211–215

[56] Fierro J, Tejuca LG (1987) Non-stoichiometric surface behaviour of $LaMO_3$ oxides as evidenced by XPS. Appl Surf Sci 27:453–457

[57] Meena R, Dhaka RS (2022) Dielectric properties and impedance spectroscopy of NASICON type $Na_3Zr_2Si_2PO_{12}$. Ceram Int 48:35150–35159

[58] Song Y, Noh TW, Lee S-I, Gaines JR (1986) Experimental study of the three-dimensional ac conductivity and dielectric constant of a conductor-insulator composite near the percolation threshold. Phys Rev B 33:904

[59] Khan R, Yuan Y, Iqbal Z, Yang J, Wang W, Ye Z, Lu J (2017) Variation of structural, optical, dielectric and magnetic properties of SnO_2 nanoparticles. J Mater Sci: Mater Electron 28:4625–4636

[60] Khan S, Khan T, Khan R, Khan M, Khattak SA, Khan G (2019) Investigation of structural, optical, electrochemical and dielectric properties of SnO_2/GO nanocomposite. J Mater Sci Mater Electron 30:10202–10210

[61] Rahman MA, Chung G-S (2013) Synthesis of PVDF-graphene nanocomposites and their properties. J Alloy Compd 581:724–730

[62] Dang ZM, Wang L, Yin Y, Zhang Q, Lei QQ (2007) Giant dielectric permittivities in functionalized carbon-nanotube/electroactive-polymer nanocomposites. Adv Mater 19:852–857

[63] Mehraj S (2012) Influence of Fe ions on the structural and dielectric properties of SnO_2 nanoparticles. Sci Adv Mater 4:1258–1267

[64] Ahmed A, Siddique MN, Ali T, Tripathi P (2018) Influence of reduced graphene oxide on structural, optical, thermal and dielectric properties of SnO_2 nanoparticles. Adv Powder Technol 29:3415–3426

[65] Bhakta N, Chakrabarti P (2019) XRD analysis, Raman, AC conductivity and dielectric properties of Co and Mn co-doped SnO_2 nanoparticles. Appl Phys A 125:1–11

[66] Kamran M, Anis-ur-Rehman M (2020) Enhanced transport properties in Ce doped cobalt ferrites nanoparticles for resistive RAM applications. J Alloy Compd 822:153583

[67] Sumi S, Rao PP, Deepa M, Koshy P (2010) Electrical conductivity and impedance spectroscopy studies of cerium based aeschynite type semiconducting oxides: $CeTiMO^6$ (M= Nb or Ta). J Appl Phys 108:063718

[68] Nasri S, Oueslati A, Chaabane I, Gargouri M (2016) AC conductivity, electric modulus analysis and electrical conduction mechanism of $RbFeP_2O_7$ ceramic compound. Ceram Int 42:14041–14048

[69] Kamran M, Anis-ur-Rehman M (2022) Resistive switching effect in RE-doped cobalt ferrite nanoparticles. Ceram Int 48:16912–16922

[70] Kumar G, Sharma S, Kotnala R, Shah J, Shirsath SE, Batoo KM, Singh M (2013) Electric, dielectric and AC electrical conductivity study of nanocrystalline cobalt substituted Mg–Mn ferrites synthesized via solution combustion technique. J Mol Struct 1051:336–344

[71] Singh R, Kumar R, Kumar A, Kumar D, Kumar M (2019) Enhanced resistive switching in graphene oxide based composite thin film for nonvolatile memory applications. Mater Res Express 6:105621

[72] Zhu S, Sun B, Ranjan S, Zhu X, Zhou G, Zhao H, Mao S, Wang S, Zhao Y, Fu G (2019) Mechanism analysis of a flexible organic memristive memory with capacitance effect and negative differential resistance state. APL Mater 7:081117

[73] Lee JH, Wu C, Sung S, An H, Kim TW (2019) Highly flexible and stable resistive switching devices based on WS2 nanosheets: poly (methylmethacrylate) nanocomposites. Sci Rep 9:19316

[74] Yen T-J, Chin A, Gritsenko V (2021) Improved device distribution in high-performance SiNx resistive random access memory via arsenic ion implantation. Nanomaterials 11:1401

[75] Bhattacharjee S, Sarkar PK, Prajapat M, Roy A (2017) Electrical reliability, multilevel data storage and mechanical stability of MoS_2-PMMA nanocomposite-based nonvolatile memory device. J Phys D Appl Phys 50:265103

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

arXiv:2407.10514v1 [math.CV] 15 Jul 2024

A NEW DIFFERENTIAL SUBORDINATION TECHNIQUE FOR A SUBCLASS OF STARLIKE FUNCTIONS

S. SIVAPRASAD KUMAR AND POOJA YADAV

Abstract. In the present investigation, we employ a new technique to find several first and second order differential subordination implications involving the following starlike class associated with a bean shaped domain:

$$\mathcal{S}^*_{\mathfrak{B}} := \left\{ f \in \mathcal{S} : \frac{zf'(z)}{f(z)} \prec \sqrt{1+\tanh z} =: \mathfrak{B}(z) \right\}.$$

Also, we give several applications stemming from our derived results.

1. Introduction

Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ and $\mathcal{H}$ be the class of functions f analytic in $\mathbb{D}$ with $f(z) = 1+a_1z+a_2z^2+\cdots$. Also, let $\mathcal{A}$ be the class of analytic functions f normalized by $f(0) = f'(0)-1 = 0$ in $\mathbb{D}$ and $\mathcal{S}$ be the subclass of $\mathcal{A}$ consisting of all univalent functions. Further $\mathcal{S}^*$ is a subclass of $\mathcal{S}$, which consists of all functions f such that $f(\mathbb{D})$ is starshaped with respect to origin and these functions are known as starlike functions with respect to origin. Analytically, $f \in \mathcal{S}^*$ if we have $\operatorname{Re} zf'(z)/f(z) > 0$ Furthermore, to compare the inclusion properties of two analytic functions f and g, having same origin, i.e. $f(0) = g(0)$, we say that f is subordinate to g, written as $f \prec g$, if and only if $f(z) = g(w(z))$ where $w(z)$ is a Schwarz function. In terms of subordination, Ma and Minda [9] unified all the subclasses of $\mathcal{S}^*$ by defining the following class:

$$\mathcal{S}^*(\phi) = \left\{ f \in \mathcal{S} : \frac{zf'(z)}{f(z)} \prec \phi(z) \right\}, \tag{1.1}$$

where ϕ is a univalent and analytic function with positive real part, starlike with respect to $\phi(0) = 1$, $\phi'(0) > 0$ and $\phi(\mathbb{D})$ is symmetric with respect to the real axis. Several well-known classes can be obtained by specializing ϕ, such as the Janowski strongly starlike class of order α $\mathcal{SS}^*(A, B, \alpha) := \mathcal{S}^*(((1+Az)/(1+Bz))^\alpha)$ where $\alpha \in (0, 1]$ and $A, B \in \mathbb{C}$ with $|B| \leq 1$ and $A \neq B$, studied by Kumar and Yadav in [7]. Here $\mathcal{SS}^*(1, -1, \alpha) =: \mathcal{SS}^*(\alpha)$; the class of strongly starlike functions of order α. And for $A, B \in [-1, 1]$ and $\alpha = 1$ we obtain the class of Janowski starlike functions, denoted as $\mathcal{S}^*(A, B)$, which specifically reduces to many well-known classes, for instance see [5]. Sokól and Stankiewicz [15] considered the class $\mathcal{S}^*_L$ by opting $\phi(z) = \sqrt{1+z}$. Geometrically $\mathcal{S}^*_L = \{f \in \mathcal{S}; |(zf'(z)/f(z))^2 - 1| < 1\}$, that is, $zf'(z)/f(z)$ lies in the domain bounded by the right lemniscate of Bernoulli. Moreover, Goel and Kumar [4] studied the starlike class associated with sigmoid, $\mathcal{S}^*(2/(1+e^{-z})) =: \mathcal{S}^*_{SG}$ and Mendiratta et al. [11] introduced and studied the class associated with exponential, $\mathcal{S}^*(e^z) =: \mathcal{S}^*_e$. Similarly, in [8] the following class was introduced by Kumar and Yadav:

$$\mathcal{S}^*_{\mathfrak{B}} := \left\{ f \in \mathcal{S} : \frac{zf'(z)}{f(z)} \prec \sqrt{1+\tanh z} =: \mathfrak{B}(z) \right\}.$$

2010 *Mathematics Subject Classification.* 30C45, 30C80.

Key words and phrases. Analytic functions, Starlike functions, Subordination, Admissibility condition.

Observe that the function $\mathfrak{B}(z) := \sqrt{1+\tanh z} = \sqrt{2/(1+e^{-2z})}$ conformally maps $\mathbb{D}$ onto the following bean shaped region

$$\Omega_{\mathfrak{B}} := \left\{ w \in \mathbb{C} : \left| \log\left(\frac{w^2}{2-w^2}\right) \right| < 2 \right\}. \tag{1.2}$$

In [8], authors obtained various geometric characteristics of $\mathfrak{B}(z)$ and derived several inclusion and radii results associated with $\mathcal{S}^*_{\mathfrak{B}}$. The present article is an extension of that work in terms of differential subordination. Here, we are establishing certain first and second order differential subordination implications related to the function $\mathfrak{B}(z)$, utilizing a lemma that significantly simplifies the calculations. Differential subordination, analogous to differential inequalities in real analysis, has numerous applications. A pivotal contribution to this field was made by Miller and Mocanu's influential monograph [12], which sparked a major revolution in the geometric function theory research community. Many researchers have addressed problems related to differential subordination, such as those in [3, 6, 13, 14, 16]. Chapter 2 of the aforementioned monograph presents a set of results that establish admissibility conditions for second-order differential subordination. Significant work has been built upon these results. For instance, Madaan et al. [10] investigated the class of admissible functions associated with the lemniscate of Bernoulli and proved several first- and second-order differential subordination implications. In [16], the author established various second- and third-order differential subordination implications by defining admissibility conditions for the exponential function. In [1], the authors modified the concept of second-order differential subordination by introducing β-admissible functions, and in [2], the authors extended the concept of differential subordination and admissibility conditions to third order.

We now present essential definitions, lemmas, and notations necessary for further discussion.

Lemma 1.1. [8] *Let $w \in \mathbb{C}$ such that $|w| > \sqrt{2}$ and $R_0 = e\sqrt{2/(e^2-1)} \simeq 1.5208$. Then*

$$\left| \log\left(\frac{w^2}{2-w^2}\right) \right| \geq 2, \quad \textit{whenever} \quad |w| \geq R_0.$$

Definition 1.2. [12] Let $\mathcal{Q}$ denotes the set of functions q that are analytic and univalent on $\overline{\mathbb{D}}\backslash\mathcal{E}(q)$, where

$$\mathcal{E}(q) = \left\{ \zeta \in \partial\mathbb{D} : \lim_{z\to\zeta} q(z) = \infty \right\}$$

such that $q'(\zeta) \neq 0$ for $\zeta \in \partial\mathbb{D}\backslash\mathcal{E}(q)$.

Definition 1.3. [12] Let Ω be a set in $\mathbb{C}$, $q \in \mathcal{Q}$ and n be a positive integer, the class of admissible functions $\Psi_n(\Omega, q)$, consists of those functions $\psi : \mathbb{C}^3 \times \mathbb{D} \to \mathbb{C}$ that satisfy the admissibility condition $\psi(r, s, t; z) \notin \Omega$, whenever $r = q(\zeta)$, $s = m\zeta q'(\zeta)$,

$$\operatorname{Re}\left(\frac{t}{s}+1\right) \geq m \operatorname{Re}\left(\frac{\zeta q''(\zeta)}{q'(\zeta)}+1\right),$$

$z \in \mathbb{D}$, $\zeta \in \partial\mathbb{D}\backslash\mathcal{E}(q)$ and $m \geq n$. In particular, we denote $\Psi_1(\Omega, q)$ as $\Psi(\Omega, q)$.

Lemma 1.4. [12] *Let $\psi \in \Psi(\Omega, q)$ with $q(0) = 1$. If $p \in \mathcal{H}$ satisfies*

$$\psi(p(z), zp'(z), z^2p''(z); z) \in \Omega,$$

then $p \prec q$.

Throughout this article, we consider the function $q(z) = \mathfrak{B}(z)$ and define the admissibility class $\Psi(\Omega, q)$, where $\Omega \subset \mathbb{C}$. Since $\mathcal{E}(\mathfrak{B}) = \varnothing$, therefore $\mathfrak{B} \in \mathcal{Q}$. For $\zeta = e^{i\theta}$ $(0 \leq \theta < 2\pi)$ we have,

$$q(\zeta) = \sqrt{\frac{2}{1+e^{-2e^{i\theta}}}}, \quad \zeta q'(\zeta) = \frac{\sqrt{2}e^{i\theta}e^{-2e^{i\theta}}}{(1+e^{-2e^{i\theta}})^{3/2}}$$

and

$$\operatorname{Re}\left(\frac{\zeta q''(\zeta)}{q'(\zeta)}\right) = \frac{\cos\theta - 2e^{4\cos\theta}\cos\theta + e^{2\cos\theta}(\cos(\theta - 2\sin\theta) - 2\cos(\theta + 2\sin\theta))}{1 + e^{4\cos\theta} + 2e^{2\cos\theta}\cos(2\sin\theta)} =: g(\theta). \quad (1.3)$$

Here,

$$|q(\zeta)| = \frac{\sqrt{2}}{(1 + e^{-4\cos\theta} + 2e^{-2\cos\theta}\cos(2\sin\theta))^{\frac{1}{4}}} =: \omega(\theta). \quad (1.4)$$

From [8], we obtain that the minimum value of $\omega(\theta)$ is $\omega(\pi) = \sqrt{2/(1+e^2)}$ and maximum is $\omega(\theta_0) =: \omega_{\theta_0} \simeq 1.438$, where $\theta_0 \simeq 1.31364$. By simple analysis we also obtain that $\min g(\theta) = g(0)$, and

$$|\zeta q'(\zeta)| = \frac{\sqrt{2}e^{-2\cos\theta}}{(1 + e^{-4\cos\theta} + 2e^{-2\cos\theta}\cos(2\sin\theta))^{3/4}} := d(\theta). \quad (1.5)$$

Further, a computation shows that $d'(\theta) = 0$ if and only if $\theta \in \{0, \theta_1, \pi\}$ where $\theta_1 \simeq 1.639$. We observe that $d(\theta)$ is increasing in $[0, \theta_1]$ and decreasing in $[\theta_1, \pi]$. Therefore, $\max d(\theta) = d(1.639) =: d_{\theta_1} \simeq 0.905$ and $\min d(\theta) = \min\{d(0), d(\pi)\} = \min\{0.158, 0.304\} = d(0)$. Moreover, $\Psi(\Omega, \mathfrak{B})$ is defined to be the class of functions $\psi : \mathbb{C}^3 \times \mathbb{D} \to \mathbb{C}$, which satisfies the following conditions:

$$\psi(r, s, t; z) \notin \Omega,$$

whenever

$$r = \mathfrak{B}(\zeta) = \sqrt{\frac{2}{1 + e^{-2e^{i\theta}}}}; \quad s = m\zeta\mathfrak{B}'(\zeta) = \frac{\sqrt{2}me^{i\theta}e^{-2e^{i\theta}}}{(1 + e^{-2e^{i\theta}})^{3/4}}; \quad \operatorname{Re}\left(1 + \frac{s}{t}\right) = m(1 + g(\theta)),$$

where $z \in \mathbb{D}$, $0 \le \theta < 2\pi$ and $m \ge 1$. Here, we start from first order differential subordination.

2. First order differential subordination

We first establish the following lemmas to prove our main results:

Lemma 2.1. *Let $-1 < B \le 0 < A \le 1$ and $w \in \mathbb{C}$ such that $|w| > 1$, then*

$$\left|\frac{w - 1}{A - Bw}\right| \ge 1, \quad \text{whenever} \quad |w| \ge \frac{1 + A}{1 + B}.$$

Proof. Let $w = Re^{i\theta}$ where $R > 1$ and $\theta \in (0, 2\pi]$. Then

$$\left|\frac{w - 1}{A - Bw}\right| \ge 1$$

is equivalent to

$$g(R, \theta) := \frac{1 + R^2 - 2R\cos\theta}{A^2 + B^2R^2 - 2ABR\cos\theta} \ge 1.$$

A simple computation shows that $g(R, \theta)$ attains its minimum at $\theta = 0$. Thus the above inequality holds if $g(R, 0) \ge 1$, i.e.

$$\frac{(R - 1)^2}{(A - BR)^2} \ge 1. \quad (2.1)$$

Since $|w| = R > 1$ and $B \le 0$ then $BR \le B < A$, thus (2.1) reduces to

$$\frac{R - 1}{A - BR} \ge 1 \implies R \ge \frac{1 + A}{1 + B}.$$

Thus the result holds. ∎

Lemma 2.2. *Let $w \in \mathbb{C}$ such that $|w| > 1$, then*

$$\left|w^2 - 1\right| \ge 1, \quad \text{whenever} \quad |w| \ge \sqrt{2}.$$

Proof. Let $w = Re^{i\theta}$ where $R > 1$ and $\theta \in (0, 2\pi]$. Then

$$\left|w^2 - 1\right| \geq 1$$

is equivalent to

$$L(R, \theta) := 1 + R^4 - 2R^2 \cos 2\theta \geq 1.$$

Clearly, the above inequality holds if $\min L(R, \theta) \geq 1$. Since, $\min L(R, \theta) = L(R, 0)$ and $R > 1$, thus the result holds. ∎

In this section, we find conditions on $\psi(p(z), zp'(z); z)$, so that the following implication holds:

$$\psi(p(z), zp'(z); z) \prec h(z) \implies p(z) \prec \mathfrak{B}(z),$$

where $h(z)$ is either $\mathfrak{B}(z)$, $(1 + Az)/(1 + Bz)$ $(-1 < B \leq 0 < A \leq 1)$ or $\sqrt{1+z}$. Here, $\psi(p(z), zp'(z); z)$ is taken to be as follows: $(p(z))^\delta + \beta(zp'(z))^n$, $(p(z))^\delta + \beta zp'(z)/(p(z))^n$ and $p(z) + zp'(z)/(\beta p(z) + \gamma)$.

Theorem 2.3. *Let n be a positive integer, $\delta \in \{0, 1\}$, $R_0 = e\sqrt{2/(e^2 - 1)}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta \in \mathbb{C}$ such that $|\beta| \geq (R_0 + \omega_{\theta_0}^\delta)(1 + e^2)^{3n/2}/(\sqrt{2}e)^n$. Suppose $p(z) \in \mathcal{H}$ and satisfies the subordination*

$$(p(z))^\delta + \beta(zp'(z))^n \prec \mathfrak{B}(z),$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1 + e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = q(\mathbb{D}) = \Omega_{\mathfrak{B}}$. Now consider $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ defined by $\psi(r, s; z) = r^\delta + \beta s^n$. For ψ to lie in the admissible class $\Psi(\Omega_{\mathfrak{B}}, \mathfrak{B})$,it must satisfy

$$\psi\left(\sqrt{\frac{2}{1 + e^{-2e^{i\theta}})}}, \frac{\sqrt{2}me^{i\theta}e^{-2e^{i\theta}}}{(1 + e^{-2e^{i\theta}})^{3/2}}; z\right) \notin \Omega_{\mathfrak{B}} \qquad (z \in \mathbb{D}, 0 \leq \theta < 2\pi, m \geq 1),$$

which is equivalent to

$$\left|\log\left(\frac{(r^\delta + \beta s^n)^2}{2 - (r^\delta + \beta s^n)^2}\right)\right| \geq 2.$$

We now use the Lemma 1.1 to prove the result, for which, let us assume $w = r^\delta + \beta s^n$. From (1.4) and (1.5) respectively, we have $|r| = |\mathfrak{B}(\zeta)| = |q(\zeta)| = \omega(\theta) \leq \max \omega(\theta) = \omega_{\theta_0}$ and $|s|/m = |\zeta\mathfrak{B}'(\zeta)| = |\zeta q'(\zeta)| = d(\theta) \geq \min d(\theta) = d(0)$. Thus

$$\begin{aligned} |w| &= |r^\delta + \beta s^n| \\ &\geq |\beta||s|^n - |r|^\delta \\ &\geq |\beta| m^n \left(d(0)\right)^n - \omega_{\theta_0}^\delta. \end{aligned}$$

Since $m \geq 1$ and $|\beta| \geq (R_0 + \omega_{\theta_0}^\delta)(1 + e^2)^{3n/2}/(\sqrt{2}e)^n$, we have

$$|w| \geq |\beta|\left(\frac{\sqrt{2}e}{(1 + e^2)^{3/2}}\right)^n - \omega_{\theta_0}^\delta \geq R_0 > \sqrt{2},$$

therefore the result follows at once by Lemma 1.1. ∎

Corollary 2.4. *Let n be a positive integer, $\delta \in \{0, 1\}$, $R_0 = e\sqrt{2/(e^2 - 1)}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta \in \mathbb{C}$ such that $|\beta| \geq (R_0 + \omega_{\theta_0}^\delta)(1 + e^2)^{3n/2}/(\sqrt{2}e)^n$. Suppose $f \in \mathcal{A}$ such that*

$$\left(\frac{zf'(z)}{f(z)}\right)^\delta + \beta\left(\frac{zf'(z)}{f(z)}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right)\right)^n \prec \mathfrak{B}(z),$$

then $f \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$(p(z))^{\delta} + \beta(zp'(z))^{n} = \left(\frac{zf'(z)}{f(z)}\right)^{\delta} + \beta\left(\frac{zf'(z)}{f(z)}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right)\right)^{n}.$$

Therefore the result holds using Theorem 2.3. ∎

Theorem 2.5. *Let n be a positive integer, $\delta \in \{0,1\}$, $R_0 = e\sqrt{2/(e^2-1)}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta \in \mathbb{C}$ such that $|\beta| \geq (R_0 + \omega_{\theta_0}^{\delta})(1+e^2)^{(3-n)/2}/(\sqrt{2}e)^{1-n}$. Suppose $p(z) \in \mathcal{H}$ and satisfies the subordination*

$$(p(z))^{\delta} + \beta\frac{zp'(z)}{(p(z))^{n}} \prec \mathfrak{B}(z),$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = q(\mathbb{D}) = \Omega_{\mathfrak{B}}$. Now consider $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be as $\psi(r,s;z) = r^{\delta} + \beta s/r^{n}$. For ψ to lie in the admissible class $\Psi(\Omega_{\mathfrak{B}}, \mathfrak{B})$, we must have

$$\psi\left(\sqrt{\frac{2}{1+e^{-2e^{i\theta}})}}, \frac{\sqrt{2}me^{i\theta}e^{-2e^{i\theta}}}{(1+e^{-2e^{i\theta}})^{3/2}}; z\right) \notin \Omega_{\mathfrak{B}} \qquad (z \in \mathbb{D}, 0 \leq \theta < 2\pi, m \geq 1),$$

which is equivalent to

$$\left|\log\left(\frac{(r^{\delta} + \beta s/r^{n})^2}{2 - (r^{\delta} + \beta s/r^{n})^2}\right)\right| \geq 2.$$

We now use the Lemma 1.1 to prove the result, for which, let us assume $w = r^{\delta} + \beta s/r^{n}$. From (1.4), we have $|r| = \omega(\theta) \leq \omega_{\theta_0}$. Thus

$$\begin{aligned} |w| &= \left|r^{\delta} + \beta\frac{s}{r^{n}}\right| \\ &\geq |\beta|\left(\frac{2^{\frac{1-n}{2}}me^{-2\cos\theta}}{(1+e^{-4\cos\theta} + 2e^{-2\cos\theta}\cos(2\sin\theta))^{\frac{3-n}{4}}}\right) - (\omega(\theta))^{\delta} \\ &\geq |\beta|m\chi_n(\theta) - (\omega(\theta))^{\delta}, \end{aligned}$$

where

$$\chi_n(\theta) := \frac{2^{\frac{1-n}{2}}e^{-2\cos\theta}}{(1+e^{-4\cos\theta} + 2e^{-2\cos\theta}\cos(2\sin\theta))^{\frac{3-n}{4}}}.$$

After a simple computation we obtain that $\min\chi_n(\theta) = \chi_n(0)$ for all $n \geq 1$. Since $m \geq 1$ and $|\beta| \geq (R_0 + \omega_{\theta_0}^{\delta})(1+e^2)^{(3-n)/2}/(\sqrt{2}e)^{1-n}$, we have

$$|w| \geq |\beta|\left(\frac{(\sqrt{2}e)^{1-n}}{(1+e^2)^{\frac{3-n}{2}}}\right) - \omega_{\theta_0}^{\delta} \geq R_0 > \sqrt{2}.$$

Therefore, the result follows immediately from Lemma 1.1. ∎

Corollary 2.6. *Let n be a positive integer, $\delta \in \{0,1\}$, $R_0 = e\sqrt{2/(e^2-1)}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta \in \mathbb{C}$ such that $|\beta| \geq (R_0 + \omega_{\theta_0}^{\delta})(1+e^2)^{(3-n)/2}/(\sqrt{2}e)^{1-n}$. Suppose f be a function in $\mathcal{A}$ such that*

$$\left(\frac{zf'(z)}{f(z)}\right)^{\delta} + \beta\left(\frac{zf'(z)}{f(z)}\right)^{1-n}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right) \prec \mathfrak{B}(z),$$

then $f \in \mathcal{S}^{*}_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$(p(z))^{\delta} + \beta\frac{zp'(z)}{(p(z))^n} = \left(\frac{zf'(z)}{f(z)}\right)^{\delta} + \beta\left(\frac{zf'(z)}{f(z)}\right)^{1-n}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right).$$

Now the result follows by an application of Theorem 2.5. ∎

Theorem 2.7. *Let $R_0 = e\sqrt{2/(e^2-1)}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta, \gamma \in \mathbb{C}$ such that $\beta \neq 0$ with $|\beta|\omega_{\theta_0} + |\gamma| \leq \sqrt{2}e/((1+e^2)^{3/2}(R_0 + \omega_{\theta_0}))$. Suppose $p(z) \in \mathcal{H}$ and satisfies the subordination*

$$p(z) + \frac{zp'(z)}{\beta p(z) + \gamma} \prec \mathfrak{B}(z),$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = q(\mathbb{D}) = \Omega_{\mathfrak{B}}$. Now consider $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be as $\psi(r, s; z) = r + s/(\beta r + \gamma)$. For ψ to lie in the admissible class $\Psi(\Omega_{\mathfrak{B}}, \mathfrak{B})$, we must have

$$\psi\left(\sqrt{\frac{2}{1+e^{-2e^{i\theta}}}}, \frac{\sqrt{2}me^{i\theta}e^{-2e^{i\theta}}}{(1+e^{-2e^{i\theta}})^{3/2}}; z\right) \notin \Omega_{\mathfrak{B}} \qquad (z \in \mathbb{D}, 0 \leq \theta < 2\pi, m \geq 1),$$

which is equivalent to

$$\left|\log\left(\frac{(r + s/(\beta r+\gamma))^2}{2 - (r + s/(\beta r+\gamma))^2}\right)\right| \geq 2.$$

We now use the Lemma 1.1 to prove the result, for which, let us assume $w = r + s/(\beta r + \gamma)$. From (1.4) and (1.5), we have $|r| = \omega(\theta) \leq \omega_{\theta_0}$ and $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$. Thus

$$\begin{aligned} |w| &= \left|r + \frac{s}{\beta r + \gamma}\right| \\ &\geq \frac{|s|}{|\beta||r| + |\gamma|} - |r| \\ &\geq \frac{md(0)}{|\beta|\omega_{\theta_0} + |\gamma|} - \omega_{\theta_0}. \end{aligned}$$

Since $m \geq 1$ and $|\beta|\omega_{\theta_0} + |\gamma| \leq \sqrt{2}e/((1+e^2)^{3/2}(R_0 + \omega_{\theta_0}))$, we have

$$|w| \geq \frac{\sqrt{2}e}{(1+e^2)^{3/2}(|\beta|\omega_{\theta_0} + |\gamma|)} - \omega_{\theta_0} \geq R_0 > \sqrt{2},$$

therefore the result follows at once by Lemma 1.1. ∎

Corollary 2.8. *Let $R_0 = e\sqrt{2/(e^2-1)}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta, \gamma \in \mathbb{C}$ such that $\beta \neq 0$ with $\omega_{\theta_0}|\beta| + |\gamma| \leq \sqrt{2}e/((1+e^2)^{3/2}(R_0 + \omega_{\theta_0}))$. Suppose f be a function in $\mathcal{A}$ such that*

$$\frac{zf'(z)}{f(z)}\left(1 + \left(\beta\frac{zf'(z)}{f(z)} + \gamma\right)^{-1}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right)\right) \prec \mathfrak{B}(z),$$

then $f \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$p(z) + \frac{zp'(z)}{\beta p(z) + \gamma} = \frac{zf'(z)}{f(z)}\left(1 + \left(\beta\frac{zf'(z)}{f(z)} + \gamma\right)^{-1}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right)\right).$$

Therefore the result holds using Theorem 2.7. ∎

Theorem 2.9. *Let $n \in \mathbb{N}$, $\delta \in \{0,1\}$, $\beta \in \mathbb{C}\backslash\{0\}$, $\omega_{\theta_0} \simeq 1.438$ and $-1 < B \leq 0 < A \leq 1$. Suppose $p \in \mathcal{H}$ such that it satisfies*

$$(p(z))^\delta + \beta(zp'(z))^n \prec \frac{1+Az}{1+Bz}$$

where

$$|\beta| \geq \left(\frac{(1+e^2)^{3/2}}{\sqrt{2}e}\right)^n \left(\frac{1+A}{1+B} + \omega_{\theta_0}^\delta\right) \tag{2.2}$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = \{w \in \mathbb{C} : |(w-1)/(A-Bw)| < 1\}$. Let $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be defined as $\psi(r,s;z) = r^\delta + \beta s^n$ where $r = \mathfrak{B}(\xi)$, $s = m\xi\mathfrak{B}'(\xi) = m\sqrt{2}\xi e^{-2\xi}/(1+e^{-2\xi})^{3/2}$ for $z \in \mathbb{D}$, $\xi \in \overline{\mathbb{D}}$ and $m \geq 1$. For ψ to be an admissible function and for $\psi \in \Psi[\Omega, \mathfrak{B}]$ we need to show $\psi(\mathbb{D}) \not\subseteq \Omega$, which is equivalent to

$$\left|\frac{r^\delta + \beta s^n - 1}{A - B(r^\delta + \beta s^n)}\right| \geq 1.$$

We now use the Lemma 2.1 to prove the result, for which, let us assume $w = r^\delta + \beta s^n$. From (1.4) and (1.5), we have $|r| = \omega(\theta) \leq \omega_{\theta_0}$ and $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$. Thus

$$\begin{aligned} |w| &= |r^\delta + \beta s^n| \\ &\geq |\beta||s|^n - |r|^\delta \\ &\geq |\beta| m^n (d(0))^n - \omega_{\theta_0}^\delta \end{aligned}$$

Since $m \geq 1$ and using (2.2), we have

$$|w| \geq |\beta| \left(\frac{\sqrt{2}e}{(1+e^2)^{3/2}}\right)^n - \omega_{\theta_0}^\delta \geq \frac{1+A}{1+B} > 1,$$

therefore the result follows at once from Lemma 2.1. ∎

Corollary 2.10. *Let $n \in \mathbb{N}$, $\delta \in \{0,1\}$ and $\beta \in \mathbb{C}\backslash\{0\}$ such that β satisfies inequality (2.2). Moreover, let $\omega_{\theta_0} \simeq 1.438$ and $-1 < B \leq 0 < A \leq 1$. Suppose $f \in \mathcal{A}$ such that*

$$\left(\frac{zf'(z)}{f(z)}\right)^\delta + \beta\left(\frac{zf'(z)}{f(z)}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right)\right)^n \prec \frac{1+Az}{1+Bz},$$

then $f \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$(p(z))^\delta + \beta(zp'(z))^n = \left(\frac{zf'(z)}{f(z)}\right)^\delta + \beta\left(\frac{zf'(z)}{f(z)}\left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right)\right)^n.$$

Now the result follows directly from Theorem 2.9. ∎

Theorem 2.11. *Let $n \in \mathbb{N}$, $\delta \in \{0,1\}$, $\beta \in \mathbb{C}\backslash\{0\}$, $\omega_{\theta_0} \simeq 1.438$ and $-1 \leq B < 0 \leq A \leq 1$. Suppose $p \in \mathcal{H}$ such that it satisfies*

$$(p(z))^\delta + \beta\frac{zp'(z)}{(p(z))^n} \prec \frac{1+Az}{1+Bz},$$

where

$$|\beta| \geq \left(\frac{(1+e^2)^{\frac{3-n}{2}}}{(\sqrt{2}e)^{1-n}}\right)\left(\frac{1+A}{1+B} + \omega_{\theta_0}^\delta\right). \tag{2.3}$$

Then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = \{w \in \mathbb{C} : |(w-1)/(A-Bw)| < 1\}$. Let $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be defined as $\psi(r,s;z) = r^\delta + \beta s/r^n$ where $r = \mathfrak{B}(\xi)$, $s = m\xi\mathfrak{B}'(\xi) = m\sqrt{2}\xi e^{-2\xi}/(1+e^{-2\xi})^{3/2}$ for $z \in \mathbb{D}$, $\xi \in \overline{\mathbb{D}}$ and $m \geq 1$. For ψ to be an admissible function and for $\psi \in \Psi[\Omega, \mathfrak{B}]$ we need to show $\psi(\mathbb{D}) \not\subseteq \Omega$, which is equivalent to

$$\left| \frac{r^\delta + \beta\frac{s}{r^n} - 1}{A - B(r^\delta + \beta\frac{s}{r^n})} \right| \geq 1.$$

We now use the Lemma 2.1 to prove the result, for which, let us assume $w = r^\delta + \beta s/r^n$. From (1.4), we have $|r| = \omega(\theta) \leq \omega_{\theta_0}$. Thus

$$\begin{aligned} |w| &= \left| r^\delta + \beta\frac{s}{r^n} \right| \\ &\geq |\beta| \left(\frac{2^{\frac{1-n}{2}} m e^{-2\cos\theta}}{(1 + e^{-4\cos\theta} + 2e^{-2\cos\theta}\cos(2\sin\theta))^{\frac{3-n}{4}}} \right) - (\omega(\theta))^\delta \\ &\geq |\beta| m \chi_n(\theta) - (\omega(\theta))^\delta, \end{aligned}$$

where

$$\chi_n(\theta) := \frac{2^{\frac{1-n}{2}} e^{-2\cos\theta}}{(1 + e^{-4\cos\theta} + 2e^{-2\cos\theta}\cos(2\sin\theta))^{\frac{3-n}{4}}}.$$

After a simple computation we obtain that $\min \chi_n(\theta) = \chi_n(0)$ for all $n \geq 1$. Since $m \geq 1$ and using (2.3), we have

$$|w| \geq |\beta| \left(\frac{(\sqrt{2}e)^{1-n}}{(1+e^2)^{\frac{3-n}{2}}} \right) - \omega_{\theta_0}^\delta \geq \frac{1+A}{1+B} > 1,$$

therefore the result follows from Lemma 2.1. ∎

Corollary 2.12. *Let $n \in \mathbb{N}$, $\delta \in \{0,1\}$ and $\beta \in \mathbb{C}\backslash\{0\}$ such that β satisfies inequality (2.3). Moreover, let $\omega_{\theta_0} \simeq 1.438$ and $-1 < B \leq 0 < A \leq 1$. Suppose f be a function in $\mathcal{A}$ such that*

$$\left(\frac{zf'(z)}{f(z)}\right)^\delta + \beta \left(\frac{zf'(z)}{f(z)}\right)^{1-n} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right) \prec \frac{1+Az}{1+Bz},$$

then $f \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$(p(z))^\delta + \beta\frac{zp'(z)}{(p(z))^n} = \left(\frac{zf'(z)}{f(z)}\right)^\delta + \beta \left(\frac{zf'(z)}{f(z)}\right)^{1-n} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right).$$

Now the result follows by an application of Theorem 2.11. ∎

Theorem 2.13. *Let $\beta, \gamma \in \mathbb{C}$ such that $\beta \neq 0$, $\omega_{\theta_0} \simeq 1.438$ and $-1 < B \leq 0 < A \leq 1$. Suppose $p \in \mathcal{H}$ such that it satisfies*

$$p(z) + \frac{zp'(z)}{\beta p(z) + \gamma} \prec \frac{1+Az}{1+Bz}$$

provided

$$|\beta|\omega_{\theta_0} + |\gamma| \leq \left(\frac{(1+e^2)^{3/2}}{\sqrt{2}e} \left(\frac{1+A}{1+B} + \omega_{\theta_0} \right) \right)^{-1} \tag{2.4}$$

where Then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = \{w \in \mathbb{C} : |(w-1)/(A-Bw)| < 1\}$. Let $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be defined as $\psi(r,s;z) = r + s/(\beta r + \gamma)$ where $r = \mathfrak{B}(\xi)$, $s = m\xi\mathfrak{B}'(\xi) = m\sqrt{2}\xi e^{-2\xi}/(1+e^{-2\xi})^{3/2}$ for $z \in \mathbb{D}$, $\xi \in \overline{\mathbb{D}}$ and $m \geq 1$. For ψ to be an admissible function and for $\psi \in \Psi[\Omega, \mathfrak{B}]$ we need to show $\psi(\mathbb{D}) \not\subseteq \Omega$, which is equivalent to

$$\left| \frac{r + \frac{s}{\beta r+\gamma}}{A - B(r + \frac{s}{\beta r+\gamma})} \right| \geq 1.$$

We now use the Lemma 2.1 to prove the result, for which, let us assume $w = r + s/(\beta r + \gamma)$. From (1.4) and (1.5), we have $|r| = \omega(\theta) \leq \omega_{\theta_0}$ and $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$. Thus

$$\begin{aligned} |w| &= \left| r + \frac{s}{\beta r + \gamma} \right| \\ &\geq \frac{|s|}{|\beta||r|+|\gamma|} - |r| \\ &\geq \frac{md(0)}{|\beta|\omega_{\theta_0} + |\gamma|} - \omega_{\theta_0}. \end{aligned}$$

Since $m \geq 1$ and using (2.4), we have

$$|w| \geq \frac{\sqrt{2}e}{(1+e^2)^{3/2}(|\beta|\omega_{\theta_0} + |\gamma|)} - \omega_{\theta_0} \geq \frac{1+A}{1+B} > 1,$$

therefore the result follows at once from Lemma 2.1. ∎

Corollary 2.14. *Let $\beta, \gamma \in \mathbb{C}$ such that $\beta \neq 0$ and satisfies inequality (2.4). Moreover, let $\omega_{\theta_0} \simeq 1.438$ and $-1 < B \leq 0 < A \leq 1$. Suppose f be a function in $\mathcal{A}$ such that*

$$\frac{zf'(z)}{f(z)} \left(1 + \left(\beta \frac{zf'(z)}{f(z)} + \gamma \right)^{-1} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)} \right) \right) \prec \frac{1+Az}{1+Bz},$$

then $f \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$p(z) + \frac{zp'(z)}{\beta p(z) + \gamma} = \frac{zf'(z)}{f(z)} \left(1 + \left(\beta \frac{zf'(z)}{f(z)} + \gamma \right)^{-1} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)} \right) \right).$$

Thus the result follows by an application of Theorem 2.13. ∎

Theorem 2.15. *Let $n \in \mathbb{N}$, $\delta \in \{0, 1\}$, $\beta \in \mathbb{C}\backslash\{0\}$ and $\omega_{\theta_0} \simeq 1.438$. Suppose $p \in \mathcal{H}$ such that it satisfies*

$$(p(z))^\delta + \beta(zp'(z)^n \prec \sqrt{1+z}$$

where

$$|\beta| \geq \left(\frac{(1+e^2)^{3/2}}{\sqrt{2}e} \right)^n \left(\sqrt{2} + \omega_{\theta_0}^\delta \right), \tag{2.5}$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = \{w \in \mathbb{C} : |w^2 - 1| < 1\}$. Let $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be defined as $\psi(r,s;z) = r^\delta + \beta s^n$ where $r = \mathfrak{B}(\xi)$, $s = m\xi\mathfrak{B}'(\xi) = m\sqrt{2}\xi e^{-2\xi}/(1+e^{-2\xi})^{3/2}$ for $z \in \mathbb{D}$, $\xi \in \overline{\mathbb{D}}$ and $m \geq 1$. For ψ to be an admissible function and for $\psi \in \Psi[\Omega, \mathfrak{B}]$ we need to show $\psi(\mathbb{D}) \not\subseteq \Omega$, which is equivalent to

$$\left| (r^\delta + \beta s^n)^2 - 1 \right| \geq 1.$$

We now use the Lemma 2.2 to prove the result, for which, let us assume $w = r^\delta + \beta s^n$. From (1.4) and (1.5), we have $|r| = \omega(\theta) \leq \omega_{\theta_0}$ and $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$. Thus

$$\begin{aligned} |w| &= |r^\delta + \beta s^n| \\ &\geq |\beta||s|^n - |r|^\delta \\ &\geq |\beta| m^n (d(0))^n - \omega_{\theta_0}^\delta. \end{aligned}$$

Since $m \geq 1$ and using (2.5), we have

$$|w| \geq |\beta| \left(\frac{\sqrt{2}e}{(1+e^2)^{3/2}} \right)^n - \omega_{\theta_0}^\delta \geq \sqrt{2} > 1,$$

therefore the result follows from Lemma 2.2. ∎

Corollary 2.16. *Let $n \in \mathbb{N}$, $\delta \in \{0, 1\}$, $\beta \in \mathbb{C}\backslash\{0\}$ such that it satisfies (2.5) and $\omega_{\theta_0} \simeq 1.438$. Further if $f \in \mathcal{A}$ and satisfies*

$$\left(\frac{zf'(z)}{f(z)} \right)^\delta + \beta \left(\frac{zf'(z)}{f(z)} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)} \right) \right)^n \prec \sqrt{1+z}$$

Then $p(z) \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$(p(z))^\delta + \beta (zp'(z))^n = \left(\frac{zf'(z)}{f(z)} \right)^\delta + \beta \left(\frac{zf'(z)}{f(z)} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)} \right) \right)^n.$$

Now the result follows from Theorem 2.15. ∎

Theorem 2.17. *Let $n \in \mathbb{N}$, $\delta \in \{0, 1\}$, $\beta \in \mathbb{C}\backslash\{0\}$ and $\omega_{\theta_0} \simeq 1.438$. Suppose $p \in \mathcal{H}$ such that it satisfies*

$$(p(z))^\delta + \beta \frac{zp'(z)}{(p(z))^n} \prec \sqrt{1+z}$$

where

$$|\beta| \geq \left(\frac{(1+e^2)^{\frac{3-n}{2}}}{(\sqrt{2}e)^{1-n}} \right) \left(\sqrt{2} + \omega_{\theta_0}^\delta \right) \tag{2.6}$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = \{w \in \mathbb{C} : |w^2 - 1| < 1\}$. Let $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be defined as $\psi(r, s; z) = r^\delta + \beta s/r^n$ where $r = \mathfrak{B}(\xi)$, $s = m\xi\mathfrak{B}'(\xi) = m\sqrt{2}\xi e^{-2\xi}/(1+e^{-2\xi})^{3/2}$ for $z \in \mathbb{D}$, $\xi \in \overline{\mathbb{D}}$ and $m \geq 1$. For ψ to be an admissible function and for $\psi \in \Psi[\Omega, \mathfrak{B}]$ we need to show $\psi(\mathbb{D}) \not\subseteq \Omega$, which is equivalent to

$$\left| \left(r^\delta + \beta \frac{s}{r^n} \right)^2 - 1 \right| \geq 1.$$

We now use the Lemma 2.2 to prove the result, for which, let us assume $w = r^\delta + \beta s/r^n$. From (1.4), we have $|r| = \omega(\theta) \leq \omega_{\theta_0}$. Thus

$$\begin{aligned} |w| &= \left| r^\delta + \beta \frac{s}{r^n} \right| \\ &\geq |\beta| \left(\frac{2^{\frac{1-n}{2}} m e^{-2\cos\theta}}{(1 + e^{-4\cos\theta} + 2e^{-2\cos\theta}\cos(2\sin\theta))^{\frac{3-n}{4}}} \right) - (\omega(\theta))^\delta \\ &\geq |\beta| m \chi_n(\theta) - (\omega(\theta))^\delta, \end{aligned}$$

where

$$\chi_n(\theta) := \frac{2^{\frac{1-n}{2}} e^{-2\cos\theta}}{(1+e^{-4\cos\theta}+2e^{-2\cos\theta}\cos(2\sin\theta))^{\frac{3-n}{4}}}.$$

After a simple computation we obtain that $\min \chi_n(\theta) = \chi_n(0)$ for all $n \geq 1$. Since $m \geq 1$ and using (2.6), we have

$$|w| \geq |\beta| \left(\frac{(\sqrt{2}e)^{1-n}}{(1+e^2)^{\frac{3-n}{2}}} \right) - \omega_{\theta_0}^{\delta} \geq \sqrt{2} > 1,$$

therefore the result follows from Lemma 2.2. ∎

Corollary 2.18. *Let $n \in \mathbb{N}$, $\delta \in \{0,1\}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta \in \mathbb{C}\backslash\{0\}$ such that β satisfies inequality (2.6). Suppose f be a function in $\mathcal{A}$ such that*

$$\left(\frac{zf'(z)}{f(z)}\right)^{\delta} + \beta \left(\frac{zf'(z)}{f(z)}\right)^{1-n} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right) \prec \sqrt{1+z},$$

then $f \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$(p(z))^{\delta} + \beta \frac{zp'(z)}{(p(z))^n} = \left(\frac{zf'(z)}{f(z)}\right)^{\delta} + \beta \left(\frac{zf'(z)}{f(z)}\right)^{1-n} \left(1 + \frac{zf''(z)}{f'(z)} - \frac{zf'(z)}{f(z)}\right).$$

Now the result follows as an application of Theorem 2.17. ∎

Theorem 2.19. *Let $\beta, \gamma \in \mathbb{C}$ such that $\beta \neq 0$ and $\omega_{\theta_0} \simeq 1.438$. Suppose $p \in \mathcal{H}$ such that it satisfies*

$$p(z) + \frac{zp'(z)}{\beta p(z) + \gamma} \prec \sqrt{1+z}$$

provided

$$|\beta|\omega_{\theta_0} + |\gamma| \leq \left(\frac{(1+e^2)^{3/2}}{\sqrt{2}e} \left(\sqrt{2} + \omega_{\theta_0}\right) \right)^{-1} \tag{2.7}$$

where Then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \sqrt{2/(1+e^{-2z})} =: \mathfrak{B}(z)$ and $\Omega = \{w \in \mathbb{C} : |w^2 - 1| < 1\}$. Let $\psi : \mathbb{C}^2 \times \mathbb{D} \to \mathbb{C}$ be defined as $\psi(r, s; z) = r + s/(\beta r + \gamma)$ where $r = \mathfrak{B}(\xi)$, $s = m\xi\mathfrak{B}'(\xi) = m\sqrt{2}\xi e^{-2\xi}/(1+e^{-2\xi})^{3/2}$ for $z \in \mathbb{D}$, $\xi \in \overline{\mathbb{D}}$ and $m \geq 1$. For ψ to be an admissible function and for $\psi \in \Psi[\Omega, \mathfrak{B}]$ we need to show $\psi(\mathbb{D}) \not\subseteq \Omega$, i.e.

$$\left| \left(r + \frac{s}{\beta r + \gamma}\right)^2 - 1 \right| \geq 1.$$

We now use the Lemma 2.2 to prove the result, for which, let us assume $w = r + s/(\beta r + \gamma)$. From (1.4) and (1.5), we have $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$ and $|r| = \omega(\theta) \leq \omega_{\theta_0}$. Thus

$$\begin{aligned} |w| &= \left| r + \frac{s}{\beta r + \gamma} \right| \\ &\geq \frac{|s|}{|\beta||r| + |\gamma|} - |r| \\ &\geq \frac{md(0)}{|\beta|\omega_{\theta_0} + |\gamma|} - \omega_{\theta_0}. \end{aligned}$$

Since $m \geq 1$ and using (2.7), we have

$$|w| \geq \frac{\sqrt{2}e}{(1+e^2)^{3/2}(|\beta|\omega_{\theta_0}+|\gamma|)} - \omega_{\theta_0} \geq \sqrt{2} > 1,$$

therefore, by Lemma 2.2, the result holds. ∎

Corollary 2.20. *Let $\omega_{\theta_0} \simeq 1.438$ and $\beta, \gamma \in \mathbb{C}$ such that $\beta \neq 0$ and satisfies inequality (2.7). Suppose f be a function in $\mathcal{A}$ such that*

$$\frac{zf'(z)}{f(z)}\left(1+\left(\beta\frac{zf'(z)}{f(z)}+\gamma\right)^{-1}\left(1+\frac{zf''(z)}{f'(z)}-\frac{zf'(z)}{f(z)}\right)\right) \prec \sqrt{1+z},$$

then $f \in \mathcal{S}^*_{\mathfrak{B}}$.

Proof. Let $p(z) = zf'(z)/f(z)$, then

$$p(z)+\frac{zp'(z)}{\beta p(z)+\gamma} = \frac{zf'(z)}{f(z)}\left(1+\left(\beta\frac{zf'(z)}{f(z)}+\gamma\right)^{-1}\left(1+\frac{zf''(z)}{f'(z)}-\frac{zf'(z)}{f(z)}\right)\right).$$

Therefore the result holds using Theorem 2.19. ∎

3. Second order differential subordination

In this section, we consider the following second order differential subordination implication

$$(p(z))^\delta + \gamma z p'(z) + \beta z^2 p''(z) \prec h(z) \implies p(z) \prec \mathfrak{B}(z).$$

Taking $h(z)$ as $\mathfrak{B}(z)$, $(1+Az)/(1+Bz)$ $(-1 < B \leq 0 < A \leq 1)$ and $\sqrt{1+z}$.

Theorem 3.1. *Let $\delta \in \{0,1\}$, $R_0 = e\sqrt{2/(e^2-1)}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta, \gamma \in \mathbb{C}\backslash\{0\}$. If $p(z) \in \mathcal{H}$ such that*

$$(p(z))^\delta + \gamma z p'(z) + \beta z^2 p''(z) \prec \mathfrak{B}(z),$$

provided

$$\sqrt{2}e(\gamma(1+e^2)^2 + \beta(1-2e^4-e^2)) \geq (R_0+\omega^\delta_{\theta_0})(e^2+1)^{7/2} \tag{3.1}$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Let $q(z) = \mathfrak{B}(z)$ and $\Omega = q(\mathbb{D}) = \Omega_{\mathfrak{B}}$. Now consider $\psi : \mathbb{C}^3 \times \mathbb{D} \to \mathbb{C}$ be as $\psi(r,s,t;z) = r^\delta + \gamma s + \beta t$. For ψ to lie in the admissible class $\Psi(\Omega_{\mathfrak{B}}, \mathfrak{B})$, we must have

$$\psi(r,s,t;z) \notin \Omega_{\mathfrak{B}},$$

whenever

$$r = \mathfrak{B}(\zeta) = \sqrt{\frac{2}{1+e^{-2e^{i\theta}}}}; \quad s = m\zeta\mathfrak{B}'(\zeta) = \frac{\sqrt{2}me^{i\theta}e^{-2e^{i\theta}}}{(1+e^{-2e^{i\theta}})^{3/4}}; \quad \mathrm{Re}\left(1+\frac{s}{t}\right) = m(1+g(\theta)),$$

where $z \in \mathbb{D}$, $0 \leq \theta < 2\pi$ and $m \geq 1$. Thus, it is sufficient to show that

$$\left|\log\left(\frac{(r^\delta+\gamma s+\beta t)^2}{2-(r^\delta+\gamma s+\beta t)^2}\right)\right| \geq 2.$$

We now use the Lemma 1.1 to prove the result, for which, let us assume $w = r^\delta + \gamma s + \beta t$. From (1.3), (1.4) and (1.5), we have $\min g(\theta) = g(0)$, $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$ and

$|r|= \omega(\theta) \leq \omega_{\theta_0}$. Thus

$$\begin{aligned}|w| &= |r^\delta + \gamma s + \beta t| \\ &\geq \gamma|s|\mathrm{Re}\left(1+\frac{\beta}{\gamma}\frac{t}{s}\right) - |r|^\delta \\ &\geq \gamma m d(\theta)\left(1+\frac{\beta}{\gamma}(mg(\theta)+m-1)\right) - (\omega(\theta))^\delta \\ &\geq m\frac{\sqrt{2}e}{(e^2+1)^{3/2}}\left(\gamma+\beta\left(m\frac{1-2e^4-e^2}{(e^2+1)^2}+m-1\right)\right) - \omega_{\theta_0}^\delta.\end{aligned}$$

Since $m \geq 1$ and using (3.1), we have

$$|w| \geq \frac{\sqrt{2}e}{(e^2+1)^{3/2}}\left(\gamma+\beta\left(\frac{1-2e^4-e^2}{(e^2+1)^2}\right)\right) - \omega_{\theta_0}^\delta \geq R_0 > \sqrt{2},$$

therefore the result follows at once by Lemma 1.1. ∎

Theorem 3.2. *Let $\delta \in \{0,1\}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta, \gamma \in \mathbb{C}\backslash\{0\}$. If $p(z) \in \mathcal{H}$ such that*

$$(p(z))^\delta + \gamma z p'(z) + \beta z^2 p''(z) \prec \frac{1+Az}{1+Bz}, \quad -1 < B \leq 0 < A \leq 1,$$

provided

$$\sqrt{2}e(\gamma(1+e^2)^2 + \beta(1-2e^4-e^2)) \geq \left(\frac{1+A}{1+B} + \omega_{\theta_0}^\delta\right)(e^2+1)^{7/2} \tag{3.2}$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Here $q(z) = \mathfrak{B}(z)$ and $h(z) = (1+Az)/(1+Bz)$, thus

$$\Omega = h(\mathbb{D}) = \left\{w \in \mathbb{C} : \left|\frac{w-1}{A-Bw}\right| < 1\right\},$$

Now consider $\psi : \mathbb{C}^3 \times \mathbb{D} \to \mathbb{C}$ be as $\psi(r,s,t;z) = r^\delta + \gamma s + \beta t$. For ψ to lie in the admissible class $\Psi(\Omega, \mathfrak{B})$, we must have

$$\psi(r,s,t;z) \notin \Omega,$$

whenever

$$r = \mathfrak{B}(\zeta) = \sqrt{\frac{2}{1+e^{-2e^{i\theta}}}}; \quad s = m\zeta\mathfrak{B}'(\zeta) = \frac{\sqrt{2}me^{i\theta}e^{-2e^{i\theta}}}{(1+e^{-2e^{i\theta}})^{3/4}}; \quad \mathrm{Re}\left(1+\frac{s}{t}\right) = m(1+g(\theta)),$$

where $z \in \mathbb{D}$, $0 \leq \theta < 2\pi$ and $m \geq 1$. i.e. it is sufficient to show that

$$\left|\frac{r^\delta + \gamma s + \beta t - 1}{A - B(r^\delta + \gamma s + \beta t)}\right| \geq 1.$$

We now use the Lemma 2.1 to prove the result, for which, let us assume $w = r^\delta + \gamma s + \beta t$. From (1.3), (1.4) and (1.5), we have $\min g(\theta) = g(0)$, $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$ and $|r|= \omega(\theta) \leq \omega_{\theta_0}$. Thus

$$\begin{aligned}|w| &= |r^\delta + \gamma s + \beta t| \\ &\geq \gamma|s|\mathrm{Re}\left(1+\frac{\beta}{\gamma}\frac{t}{s}\right) - |r|^\delta \\ &\geq \gamma m d(\theta)\left(1+\frac{\beta}{\gamma}(mg(\theta)+m-1)\right) - (\omega(\theta))^\delta \\ &\geq m\frac{\sqrt{2}e}{(e^2+1)^{3/2}}\left(\gamma+\beta\left(m\frac{1-2e^4-e^2}{(e^2+1)^2}+m-1\right)\right) - \omega_{\theta_0}^\delta.\end{aligned}$$

Since $m \geq 1$ and by (3.2), we have

$$|w| \geq \frac{\sqrt{2}e}{(e^2+1)^{3/2}}\left(\gamma+\beta\left(\frac{1-2e^4-e^2}{(e^2+1)^2}\right)\right)-\omega_{\theta_0}^{\delta} \geq \frac{1+A}{1+B} > 1,$$

therefore, by Lemma 2.1, the result holds. ∎

Theorem 3.3. *Let $\delta \in \{0,1\}$, $\omega_{\theta_0} \simeq 1.438$ and $\beta, \gamma \in \mathbb{C}\backslash\{0\}$. If $p(z) \in \mathcal{H}$ such that*

$$(p(z))^{\delta} + \gamma z p'(z) + \beta z^2 p''(z) \prec \sqrt{1+z},$$

provided

$$\sqrt{2}e(\gamma(1+e^2)^2+\beta(1-2e^4-e^2)) \geq \left(\sqrt{2}+\omega_{\theta_0}^{\delta}\right)(e^2+1)^{7/2} \tag{3.3}$$

then $p(z) \prec \mathfrak{B}(z)$.

Proof. Here $q(z) = \mathfrak{B}(z)$ and $h(z) = \sqrt{1+z}$, thus

$$\Omega = h(\mathbb{D}) = \left\{w \in \mathbb{C} : \left|w^2-1\right| < 1\right\},$$

Now consider $\psi : \mathbb{C}^3 \times \mathbb{D} \to \mathbb{C}$ be as $\psi(r,s,t;z) = r^{\delta} + \gamma s + \beta t$. For ψ to lie in the admissible class $\Psi(\Omega, \mathfrak{B})$, we must have

$$\psi(r,s,t;z) \notin \Omega,$$

whenever

$$r = \mathfrak{B}(\zeta) = \sqrt{\frac{2}{1+e^{-2e^{i\theta}}}}; \quad s = m\zeta\mathfrak{B}'(\zeta) = \frac{\sqrt{2}me^{i\theta}e^{-2e^{i\theta}}}{(1+e^{-2e^{i\theta}})^{3/4}}; \quad \mathrm{Re}\left(1+\frac{s}{t}\right) = m(1+g(\theta)),$$

where $z \in \mathbb{D}$, $0 \leq \theta < 2\pi$ and $m \geq 1$. i.e. it is sufficient to show that

$$\left|\left(r^{\delta}+\gamma s+\beta t\right)^2 - 1\right| \geq 1.$$

We now use the Lemma 2.2 to prove the result, for which, let us assume $w = r^{\delta} + \gamma s + \beta t$. From (1.3), (1.4) and (1.5), we have $\min g(\theta) = g(0)$, $|s|/m = d(\theta) \geq \min d(\theta) = d(0)$ and $|r| = \omega(\theta) \leq \omega_{\theta_0}$. Thus

$$\begin{aligned}
|w| &= |r^{\delta} + \gamma s + \beta t| \\
&\geq \gamma|s|\mathrm{Re}\left(1+\frac{\beta}{\gamma}\frac{t}{s}\right) - |r|^{\delta} \\
&\geq \gamma m d(\theta)\left(1+\frac{\beta}{\gamma}(mg(\theta)+m-1)\right) - (\omega(\theta))^{\delta} \\
&\geq m\frac{\sqrt{2}e}{(e^2+1)^{3/2}}\left(\gamma+\beta\left(m\frac{1-2e^4-e^2}{(e^2+1)^2}+m-1\right)\right) - \omega_{\theta_0}^{\delta}.
\end{aligned}$$

Since $m \geq 1$ and by (3.3), we have

$$|w| \geq \frac{\sqrt{2}e}{(e^2+1)^{3/2}}\left(\gamma+\beta\left(\frac{1-2e^4-e^2}{(e^2+1)^2}\right)\right)-\omega_{\theta_0}^{\delta} \geq \sqrt{2} > 1,$$

therefore, by Lemma 2.2, the result holds. ∎

Declarations

Ethical Approval Not Applicable
Conflicts of interest The authors declare that they have no competing interests.
Authors' contributions All authors contributed equally.
Funding Not Applicable
Availability of data and materials Not Applicable

References

[1] R.M. Ali, S. Nagpal, V. Ravichandran, Second-order differential subordination for analytic functions with fixed initial coefficient. Bull. Malays. Math. Sci. Soc. **34** (2011), 611–629.

[2] J. Antonino and S. S. Miller, Third-order differential inequalities and subordinations in the complex plane, Complex Var. Elliptic Equ. **56** (2011), no. 5, 439–454.

[3] N.E. Cho, S. Kumar, V. Kumar and V. Ravichandran, Differential subordination and radius estimates for starlike functions associated with the Booth lemniscate. Turk. J. Math. **42** (2018), no. 3, 1380–1399.

[4] P. Goel and S. S. Kumar, Certain class of starlike functions associated with modified sigmoid function, Bull. Malays. Math. Sci. Soc. (2019). https://doi.org/10.1007/s40840-019-00784-y.

[5] W. Janowski, Extremal problems for a family of functions with positive real part and for some related families, Ann. Polon. Math. **23** (1970/71), 159–177.

[6] S. S. Kumar and P. Goel, Starlike functions and higher order differential subordinations, Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Mat. RACSAM **114** (2020), no. 4, Paper No. 192, 23 pp.

[7] S. S. Kumar and P. Yadav, On Oblique Domains of Janowski Function, Math. Slovaca **73** (2023), no. 2, 407–424.

[8] S. S. Kumar and P. Yadav, On starlike functions associated with the bean shaped domain, arXiv preprint arXiv:2403.14162, (2024).

[9] W. C. Ma and D. Minda, A unified treatment of some special classes of univalent functions, in *Proceedings of the Conference on Complex Analysis (Tianjin, 1992)*, 157–169, Conf. Proc. Lecture Notes Anal., I, Int. Press, Cambridge, MA.

[10] V. Madaan, A. Kumar and V. Ravichandran, Starlikeness associated with lemniscate of Bernoulli. Filomat **33** (2019), no. 7, 1937–1955.

[11] R. Mendiratta, S. Nagpal and V. Ravichandran, On a subclass of strongly starlike functions associated with exponential function, Bull. Malays. Math. Sci. Soc. **38** (2015), no. 1, 365–386.

[12] S. S. Miller and P. T. Mocanu, *Differential subordinations*, Monographs and Textbooks in Pure and Applied Mathematics, **225** (2000), Marcel Dekker, Inc., New York.

[13] S.S. Miller and P.T. Mocanu, Subordinants of differential superordinations. Complex Var. Theory Appl. **48** (2003), no. 10, 815–826.

[14] A. Saliu, K. Jabeen and V. Ravichandran, Differential subordination for certain strongly starlike functions, Rendiconti del Circolo Matematico di Palermo Series, **2** (2023), 1–18.

[15] J. Sokół and J. Stankiewicz, Radius of convexity of some subclasses of strongly starlike functions, Zeszyty Nauk. Politech. Rzeszowskiej Mat. No. 19 (1996), 101–105.

[16] N. Verma and S. S. Kumar, Differential subordination of certain class of starlike functions, Complex Var. Elliptic Equ. (2024), 1–19. doi: 10.1080/17476933.2024.2326833.

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
Email address: spkumar@dce.ac.in

Department of Applied Mathematics, Delhi Technological University, Delhi–110042, India
Email address: poojayv100@gmail.com

Electrical Engineering
https://doi.org/10.1007/s00202-024-02577-4

ORIGINAL PAPER

A novel hybrid intelligent approach for solar photovoltaic power prediction considering UV index and cloud cover

Rahma Aman[1] · M. Rizwan[1] · Astitva Kumar[2]

Received: 13 August 2023 / Accepted: 26 June 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
The power generation from photovoltaic plants depends on varying meteorological conditions. These meteorological conditions such as solar irradiance, temperature, and wind speed are nonlinear and stochastic, thus affecting the estimation of solar photovoltaic (PV) power. Accurate estimation of photovoltaic power is essential for enhancing the functioning of solar power installations. The paper aims to develop a novel deep learning-based photovoltaic power prediction model on different weather conditions. The proposed model utilises a two-stage deep learning framework for accurate solar PV power prediction, which combines the long short-term memory (LSTM) and convolutional neural network (CNN) deep learning architectures. The key role of CNN layer is to identify the weather conditions, i.e., sunny, partially cloudy, and extremely cloudy, while the LSTM layer learns the patterns of solar power generation that depend on weather variations to estimate photovoltaic power. The proposed hybrid models consider meteorological factors, such as wind speed, irradiance, temperature, and humidity, including cloud cover and UV index to provide precise solar PV power prediction. The presented hybrid model has better performance metrices having root mean square error of 0.0254, 0.03465, and 0.0824, mean square error of 0.000645, 0.00120, and 0.00679, R^2 of 0.9898, 0.9872, and 0.9358, and mean average error of 0.0163, 0.0236, and 0.2521 for sunny, partially cloudy, and extremely cloudy weather conditions, respectively. The results demonstrate that presented deep learning-based novel solar PV power prediction model can accurately predict solar PV power based on instantaneous changes in generated power patterns and aid in the optimisation of PV power plant operations. The paper presents an effective methodology for prediction of solar power that can contribute to the improvement of solar power generation and management.

Keywords Renewable energy · Photovoltaic · Deep learning · Artificial intelligence · Meteorological parameters

1 Introduction

In response to environmental concerns about greenhouse gas emissions and fossil fuel finiteness, renewable energy sources including solar, wind, hydropower, geothermal, and biomass have grown in popularity over the past several decades. Renewable energy sources are a sustainable alternative to fossil fuels due to their low carbon emissions and environmental effect, according to science. PV power is widely regarded as one of the most promising forms of renewable energy that has the potential to meet the requirements for energy production on a global scale [1]. According to reports, the renewable energy sector experienced a challenging phase in the first half of 2020, coinciding with the start of the COVID-19 epidemic. Nonetheless, the PV industry achieved a new record for new capacity in the second half of 2020, bringing the solar sector to its greatest level of more than 800 Gigawatts. Nevertheless, because solar energy is stochastic and unpredictable, it has created a variety of issues, such as voltage variations and uncertainties in a power grid, making it difficult to maintain a balance between power generation and load demand [2]. The fluctuation of solar irradiance received by PV modules demonstrates significant variability on days with cloud cover, mostly due to the dynamic

✉ Rahma Aman
rahmaaman96@gmail.com

M. Rizwan
rizwan@dce.ac.in

Astitva Kumar
astitva.kumar@nsut.ac.in

[1] Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India

[2] Department of Electrical Engineering, Netaji Subhas University of Technology, Delhi, India

Published online: 09 July 2024

movement of clouds. The presence of stochastic variations significantly influences the production of solar PV systems. Given the non-stationary and nonlinear characteristics of this phenomenon, it is crucial to accurately forecast solar irradiance. The employment of accurate predictions is crucial in improving the dependability of photovoltaic plants and enabling efficient management of supply and demand in the realm of solar energy utilisation. There are multiple approaches available for predicting solar radiation, each designed to account for the nuances and complexity inherent in this dynamic process [3, 4]. Similarly, the demand for solar PV power forecasting arises from the inherent dynamism and intermittent characteristics of solar energy. Precise predictions facilitate the dependable and effective incorporation of solar energy into the wider energy framework, yielding advantages for grid management, energy strategizing, market transactions, and sustainability endeavours as a whole.

Due to the rising prevalence of solar PV systems in built environments, accurate forecasting is simultaneously becoming more essential and more difficult to achieve.

Accurate forecasting requires two critical components: minimising the negative impact of random solar PV power on the power system and giving and predicting solar PV power output data to grid operators. As a result, anticipating the output power of solar systems is essential for the effective operation of the electrical grid. The estimation of reserves is strongly reliant on the efficient administration and scheduling of a power system. It should be emphasised that solar forecasting is becoming increasingly relevant owing to the worldwide expansion in solar PV power output [5]. For this reason, the researchers looked at a number of different methods for predicting solar PV power. Statistical techniques (ST), artificial intelligence techniques (AI), and hybrid method techniques (HM) are the major categories that these approaches fall under [6]. Interdependence and correlation studies examined how climatic conditions affect performance metrics. Some variables are interdependent, but separating their correlation effects is challenging [7].

Predicting electricity generation from meteorological factors using random forest approach and pre-processing methods like feature selection and principal component analysis, forecasting models were assessed for computation time, accuracy, and statistical indicators [8]. In order to forecast the power output of a 960 kW_P grid-connected photovoltaic (PV) plant that was erected in Italy, statistical approaches that were developed on the basis of multiregression analysis and the Elmann artificial neural network (ANN) were used. Several inputs into the artificial neural network (ANN) included time series of both the PV power that was generated and the meteorological factors that were monitored [9]. This study investigates time series statistical, physical, ensemble, machine, and deep learning solar PV power forecasting approaches. Moreover, model parameter optimisation processes, major factors influencing solar PV power predictions, and input selection for solar PV power generation forecasting models are investigated. This research compares deterministic and probabilistic forecasting approaches to help solar PV power forecasting meet current grid system issues [10]. Authors in [11] proposed a simple yet accurate short-term power forecasting method for large-scale grid-connected photovoltaic plant (LS-GCPVs.) Three artificial neural network (ANN) models for bright, partly cloudy, and overcast days are developed using a 1-year record of solar irradiance, cell temperature, and power production from a 1-MWp photovoltaic facility in Southern Italy. Some statistical methods used for solar prediction, in [12, 13] statistical models, on the other hand, can only learn from linear data since they are incapable of processing complex information. As a result, ST-based techniques are not recommended for tackling problems that need nonlinear predictions, such as those related to solar PV power. It has become clear that AI-based models perform far better than their physical and statistical counterparts. It has been shown that AI-based models are more effective than physical and statistical ones. Author in [14] proposed incident solar radiation using limited meteorological dataset, and a wavelet-coupled support vector machine (W-SVM) model was utilised to forecast global incident solar radiation. The predictor variables for this model included sunshine hours (St), minimum temperature (T_{min}), maximum temperature (T_{max}), wind speed (U), evaporation (E), and precipitation (P). Author in [15] uses wavelet transform (WT) and artificial intelligence (AI) to anticipate a PV system's one-hour power production using solar radiation and temperature data. The suggested approach uses WT to improve PV power time series data and AI to better capture nonlinear PV variability. Modern machine learning methods are used to increase PV system dependability and stability. Time series long short-term memory (LSTM) network, ConvLSTM, CNN, RF, SVM, and XGBoost regression models predicted Johannesburg's sun irradiance [16]. AI-based approaches outperform traditional ones. AI-based approaches can handle imperfect inputs, update and maintain easily, and reason. A deep learning model using a group of gated recurrent units (GRU) and optimised hyperparameters uses sub-signals from similar days as network inputs. The solar PV power forecasting is calculated by integrating each sub-signal's projection. Author in [17] develops the random forest combines present and previous solar PV power predictions from different models and meteorological data to enhance day-ahead solar PV power estimates. The year-long performance of the combined model is compared to different combining methods. Smart persistence, artificial neural network, and random forest are compared to forecast global horizontal, beam normal, and diffuse horizontal sun irradiance in Odeillo, France, a place with significant

weather variability [18]. Author in [19] proposes a five-layer CNN-LSTM model for solar PV power projections using data from Temixco, Morelos, Mexico. In the hybrid model, the convolutional layer filters local data features, while the long short-term memory network extracts temporal characteristics. Finally, the five-layer hybrid model is compared against a single LSTM, a two-layer CNN-LSTM hybrid model, and two prominent benchmarks. Author in [20] introduces hybrid deep learning forecasting. Levy-flight beetle antennae search technique is used to find prior days comparable to the projected day from real-time large data.

CNNs and LSTM networks dominate deep learning. Time series data are used for such models because LSTM networks can capture sequence pattern information and CNN models can extract important features and filter out noise. CNNs can extract patterns of local trend and the same pattern that appears in different regions of time series data, but they are not usually adapted for long temporal dependencies. LSTM networks, on the other hand, are designed to work with temporal correlations but only use the attributes provided in the training set. Hence, a hybrid model using both deep learning techniques may increase forecasting accuracy. Hybrid model forecasting predicts power effectively.

The authors in [21] proposed a CNN-LSTM model to predict irregularities in PV power generation (PVPG) that other machine learning models could not learn well. The hybrid model can accurately predict how much solar PV power will be made by taking features from the things that affect how much solar PV power will be made. Then, to show how useful the model was, they compared it to other machine learning methods. From multivariate time series, the model learns trends in the PVPG and information about how time changes over time. The results show that the 5D CNN-LSTM model predicts PVPG very well and does better than the single LSTM model and the 2D CNN-LSTM model. In [22] day-ahead solar photovoltaic energy forecasting is performed using three different long short-term memory (LSTM) networks: LSTM, bidirectional long short-term memory (BiLSTM), and stacked LSTM, taking into account the power of the PV panels, the solar energy value produced, and the weather-related features. Chen et al. [23] proposed a photovoltaic power forecasting (PVPF) model that predicts the next day's production by grouping days by their weather conditions (sunny, cloudy, or rainy) and weather variables (solar radiation, total, and low cloud coverage). Accurate forecasting of solar radiation is required for predicting the output power of a photovoltaic power plant. This article assesses the possibility of a 20 MW solar photovoltaic power plant in Zahedan for this purpose. MLP, LSTM, GRU, CNN, and CNN-LSTM models are used to forecast solar radiation and temperature for the next 10 years (1984-2021) using monthly data. CNN outperforms other models with four input parameters: global horizontal irradiance, temperature, surface pressure, relative humidity (RH), and two output parameters: temperature and radiation [24]. To address the intermittent and stochastic character of photovoltaic (PV) power generation, [25] provides an ultra-short-term prediction technique employing a convolutional neural network (CNN) and long short-term memory (LSTM) hybrid model powered by empirical wavelet transform (EWT). An LSTM-ARMA combination model driven by ensemble empirical mode decomposition (EEMD) is suggested for ultra-short-term PV output prediction to minimise solar PV power generation intermittency and uncertainty [26]. In advanced concentrated solar power system management, short-term forecasting of direct normal irradiance (DNI), photons that did not interact with the environment on their approach to the observer, is important. Hybrid model for intra-hour DNI forecasting (5, 10, and 15 min). This hybrid technique uses a knowledge-based methodology for clear-sky DNI predictions using DNI data and a machine learning algorithm [27]. As the primary input, the proposed model employs a series of time-dependent irradiance estimates near the target location. These estimates are derived from satellite images and combined with other secondary inputs in an advanced neural network with convolutional and dense layers that has been trained using a deep learning technique [28].

The construction of renewable energy power plants is currently underway, resulting in a significant increase in energy production. The prediction of power output holds significant importance from an economic standpoint. Moreover, the act of predicting solar PV power output plays a crucial role in the strategic organisation of power production, enabling effective mitigation of challenges related to system stability and power-generating equilibrium. The field of renewable energy power production forecasting has been widely investigated using various methodologies, including physical, statistical, machine learning (ML), and artificial neural network (ANN) approaches.

Table 1 represents work considering artificial intelligence model models like ANN and deep learning LSTM and CNN are most accurate model for predicting irradiance, most of the authors have considered input parameters like wind speed, temperature, irradiance, and humidity for training the model for accurate prediction very few works is been done considering the cloud cover parameters and UV index for precise prediction. As reviewed, the above combination of LSTM and CNN models is best for precise solar PV power prediction.

Motivated by the above-discussed papers, we propose a CNN-LSTM model that tries to fully use CNN's superiority for extracting spatial features and the LSTM network's superiority for extracting time series features in order to increase the performance. The model that uses both temporal and spatial for training the deep learning models is proposed.

Table 1 Solar PV power prediction and irradiance prediction using deep learning approach

Ref	Method Proposed	Input Parameters	Dataset	Performance Metric	Observation
[29]	LSTM-CNN	Relative humidity, temperature, pressure, global horizontal irradiance (GHI), wind speed, and cloud type	1 Year	27.38 Wm − 2 – 37.02 Wm − 2 (MAE)	In several seasons across a year and under three distinct sky conditions, the suggested model outperforms the standard model in predicting accuracy
[30]	LSTM & ANN	Temperature, humidity, cloudiness, radiation, and two seasonal months (month of year and day of month)	1 Year	LSTM;1.23–1.82% (RMSE) ANN; 1.67–8.02% (RMSE)	When the offered methods were compared, LSTM outperformed ANN (single and multi-layer)
[31]	LSTM-PVPF	Wind speed, ambient temperature, global horizontal radiation, wind direction, air pressure, and daily rainfall	3 years	6 ~ 9 (MAPE)	The suggested models were less computationally costly and proved consistency, excellent accuracy, feasibility, and applicability for applications requiring large dataset sets
[32]	CLSTM	Solar irradiance	12 years	≈1.515% (RMSE) ≈4.672% (MAE) ≈1.233% (APB)	The suggested model outperformed other competing models in terms of percentage error; it had higher prediction performance and outperformed standalone models such as CNN, LSTM, and DNN
[33]	DWT-CNN-LSTM	Solar irradiance	8 years	32.411% (RMSE)	The enhanced forecasting accuracy of the proposed method via a comparison with traditional and single DL models
[34]	CNN–LSTM	Irradiation, Wind speed Temperature	6 years	5.1% (RMSE) 1 day 29.7% (RMSE) 1 h 10% (RMSE) 1 week	CNN-simple and CNN-LSTM perform best for 1-h, 1-day, and 1-week forecasts, with CNN-LSTM sometimes outperforming
[35]	CNN–LSTM	Relative humidity, GHI, DHI Wind direction, temperature	4 years	34.3% (RMSE)	This paper's deep learning model for solar power prediction shows that deep learning improves accuracy
[36]	GRU model LSTM model	GHI, Solar zenith angle, Relative humidity And Air Temperature	10 years	66% (RMSE) 67% (RMSE)	GRU is compared to long short-term memory (LSTM) utilising historical sun irradiance data and exogenous meteorological factors in GRU and LSTM considerably increased predicting accuracy, outperforming univariate and statistical models
[37]	LSTM-RNN	Humidity, Temperature, Wind speed, Precipitation - Cloud cover, Wind direction	14 years	60.139%(RMSE)	The suggested method is more accurate and saves 2% more energy than Feed forward neural network

The proposed model predicts the solar PV power output by looking at both historical PV power output data further the prediction based on local weather data, i.e., sunny, partially cloudy, and extremely cloudy which is considered in very few research papers. The temporal data show the long-term patterns and trends in the PV power output, while the local meteorological data show how the output changes over short periods of time. Generally, univariant time series prediction is done where mostly solar irradiance and solar PV power are forecasted here; in this paper, multi-inputs are used in intra-hour prediction solar PV power by using CNN-LSTM hybrid models. Many papers found have trained the models with hourly dataset which is very common in the literature review; intra-hour solar PV power prediction is found in very few papers; this paper uses 10-minute dataset to train the prediction models. Further according to the literature review, cloud cover and UV index are incorporated as inputs for predicting solar PV power in very few articles since they are an essential element that affects solar PV power prediction, the cloud cover and UV index have been modelled and then used as an input to train the deep learning models, and these parameters are included with common meteorological inputs for prediction the solar PV power. Further comparison of the proposed hybrid model is done with traditional ARIMA model and deep learning models, i.e., CNN and LSTM are done based on statistical indicators, i.e., RMSE, MSE, R^2, and MAE are done.

Main contribution of this work is as follows.

This research aims in design and implementation of a new hybrid deep learning model, i.e., CNN-LSTM for predicting solar PV power output that uses both temporal and local weather data at every 10 min aiming to enhance accuracy by capturing spatiotemporal dependencies.

The proposed work helps in improving the prediction accuracy of solar PV power, by considering the meteorological parameters such as irradiance, temperature, humidity, wind speed, cloud cover, and UV index, where the cloud cover and UV index play a significant role in the development of efficient model.

The developed model is validated for various seasons and weather scenarios such as sunny, partially cloudy, and extremely cloudy, and the performance found more accurate as compared to the existing models, i.e., LSTM and CNN.

The remaining paper is organised in five sections: Sect. 1 is the literature survey which helps in finding the gaps. In Sect. 2, the proposed hybrid CNN-LSTM model is introduced and the overview of the proposed analytical method is discussed, in Sect. 3 data preparation is given, Sect. 4 discusses the results, and Sect. 5 concludes the paper.

2 Proposed methods

2.1 CNN for feature extraction

CNN, a frequently used deep learning system for image, text, and signal inputs, comprises layers that extract object attributes. Fig. 1 displays the core construction of a CNN. In CNNs, feature extraction comprises a series of convolutional, activation, and pooling processes that gradually convert raw input data into hierarchical and useful features, allowing the network to learn and generalise from complicated patterns in the data.

The performance of models is determined by the number of stacked layers as well as the kind and size of the kernel. Convolution and pooling layers are utilised to extract deep features from the input layer data. These attributes are then passed into the fully linked layer, which categorises the result values. A CNN is used to extract hierarchical picture characteristics. Thus, a CNN is able to extract relevant information from sequential and two-dimensional input data [38]. In essence, a CNN aims to build several filters capable of extracting hidden characteristics via layer-by-layer convolution and data pooling as shown in Fig. 1. Finally, the fully connected layer merges these abstract features, and an activation function is employed to solve the classification or regression problem. The previous layer's feature maps are convolved with a convolution kernel in the convolution layer, and the output feature maps are generated using an activation function [39]. The previous layer's feature maps are convolved using a convolution kernel in the convolution layer, and the output feature maps are generated. 1D-CNNs are often used to process natural language or time series since they can handle sequential data. 1D-CNNs are different from 2D-CNNs because both the kernel that does the convolution and the sequence of data used have a one-dimensional structure. As seen in Fig. 2, the kernel of a 1D-CNN moves along a single dimension.

2.2 LSTM Model

The LSTM network is a type of recurrent network. Unlike "regular" MLP, RNNs use the input data's temporal context. The LSTM network is made up of layers that allow for the modelling of long-term dependencies in sequential data. The network has been trained to recognise patterns and trends in solar PV power, including temporal dependencies that lead to accurate forecasts. The LSTM output provides insights into the temporal dynamics of solar PV power production. This is possible in an RNN because the neurons are connected to each other in the same way more than once. An LSTM is able to store and recall information because of how the neurons in a memory cell are set up. These memory cells can hold information for an indefinite amount of time. As shown in

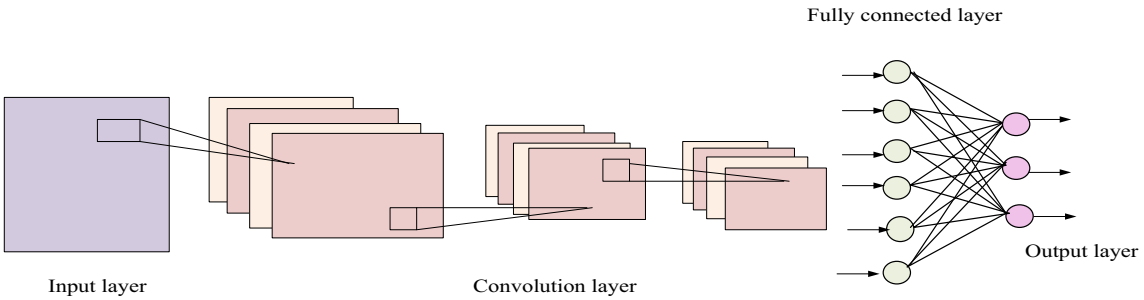


Fig. 1 Convolutional neural networks architecture

Fig. 2. 1D-CNN computation method

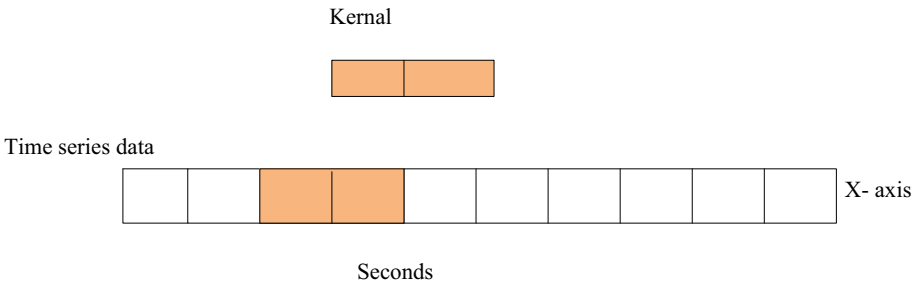


Fig. 3 Cell diagram of LSTM

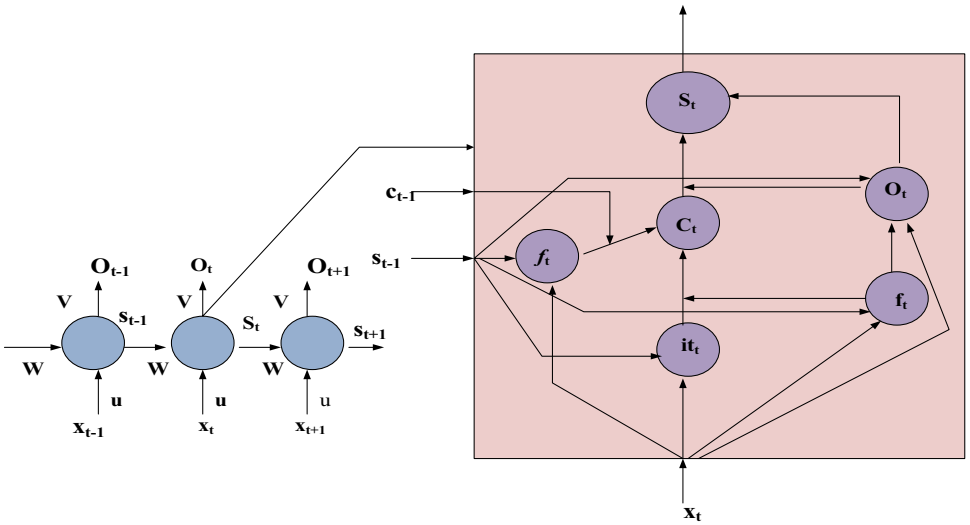


Fig. 3, the memory unit is made up of three gates labelled "Input gate (i_t)," "Forget gate (f_t)," and "Output gate (o_t)," as well as a recurrent connection. It controls how information comes into and goes out of the cell. Each gate in the LSTM gets information from the input neuron. Also, each gate has a function for turning it on and off. It has a single input, x_t, and two feedbacks from the previous state, s_{t-1} and c_{t-1}. The sigmoid activation function (g) is used by gates, while the tanh function is used by states. The LSTM's memory unit can be explained by a set of equations, where w is the weight parameter and b is the bias. An LSTM has a general structure of three layers: forget, input, and output. First, x_t and s_{t-1} are put into the LSTM architecture. Then, a decision is made about whether or not to keep the information. All of the tasks are done by the forget layer f_t [40].

$$f_t = g\left(w_{xf}\, x_t + w_{hf} s_{t-1} + b_f\right) \tag{1}$$

where activation function (g). Also, the input layer i_t is given by Eq. 2.

$$i_t = g(w_{xi} x_t + w_{hi} s_{t-1} + b_i) \tag{2}$$

$$c_t = f_t c_{t-1} + i_t i t_t \tag{3}$$

$$i t_t = \tanh(w_{xc} x_t + w_{hc} s_{t-1} + b_i t_i \tag{4}$$

Table 2 Hybrid CNN-LSTM Model parameters

Model layers		
Inputs	Size	3,7
Conv1D	Kernal Size	2
Max pooling	Pooling size	2
Conv1D	Kernal size	2
Max pooling	Pooling size	2
Flatten	–	–
LSTM layers	Hidden Neurons/Activation	64/ Relu
LSTM layers	Hidden Neurons/Activation	64/ Relu
Learning rate		0.0001
No. of epochs		500
Optimiser		Adam

The last step is to use the following expressions in the output layer to get the output data.

$$o_t = g(w_{xo}x_t + w_{ho}s_{t-1} + b_0) \tag{5}$$

$$s_t = o_t \tanh(c_t) \tag{6}$$

The process described above keeps going back and forth. This model learns weight parameters (w) and bias parameters (b) to minimise the difference between actual training values and LSTM output values.

2.3 CNN-LSTM model

In this method CNN and LSTM layers are combined as shown in Fig. 4; here, the selected inputs are sent to the CNN layer first for feature extraction, then to the flatten layer, and finally to the LSTM layer for prediction. After pre-processing, the CNN layer uses time series to find local features. The one-dimensional CNN works well for time series applications because the convolution kernel moves in a clear direction. This allows it to automatically pull out the unobserved data characteristics in the time direction. The retrieved features from the encoder (CNN) are fed into the LSTM architecture. The training data, as well as the various gates of the LSTM network, are constantly changed so that the LSTM model may find the relationships between the input and output sequence [41]. The proposed model parameters considered are shown in Table 2, and the proposed work is shown in Fig. 5.

Following are the steps for proposed work.

Stage 1—Input dimension reconstruction.

1. Inputs are selected that are correlated with the solar PV power six inputs are considered for training the models.
2. The reconstruction of inputs such as cloud cover, UV index, wind speed, irradiance, temperature, and humidity is done to extract spatial features, while historical solar PV power to obtain temporal features.
3. These inputs are given in the proposed model CNN-LSTM and LSTM and CNN deep learning model

Stage 2—Model development.

1. The deep learning models are developed with the help of Jupyter notebook platform.
2. The three deep learning models are trained individually by giving the inputs
3. Fine-tuning of parameters is done in this stage to make the model more accurate to predict solar PV power

Step 3—Evaluation and assessment.

The three deep learning models are employed for evaluating the performance in terms of three weather conditions Sunny, partially cloudy, and extremely cloudy and four different seasons (winter, summer, rainy, and post monsoon)

Stage 4—Interpret Results.

The best solar PV power prediction among the three deep learning models is analysed based on the performance metrices such as RMSE, MSE, MAE, and R^2.

3 Data preparation

This study uses dataset where irradiance and temperature are obtained from solar analyser 9018 T, humidity and wind speed are obtained from the solcast site, and cloud cover, UV index, and solar PV power are mathematically modelled for obtaining higher accuracy results. Solcast has access high-resolution historical time series and TMY data with minimal uncertainty for energy simulations and reliable solar resource assessments. Prepared for integration through API and the solar analyser is a portable instrument utilised for troubleshooting, monitoring, measuring, and analysing a variety of solar system parameters (Fig. 6). The data used are for one year from 24 November 2021 to 24 November 2022 containing 52,560 data, i.e., $365 \times 24 \times 6$, the data collected are at 10-min interval, and the dataset comprises inputs having irradiance, humidity, wind speed, cloud cover, temperatures, solar PV power, and UV index as shown in Fig. 7.

Figure 8 is a correlation matrix from where the correlation of several meteorological parameters with power has been shown, and in this proposed deep learning model from the

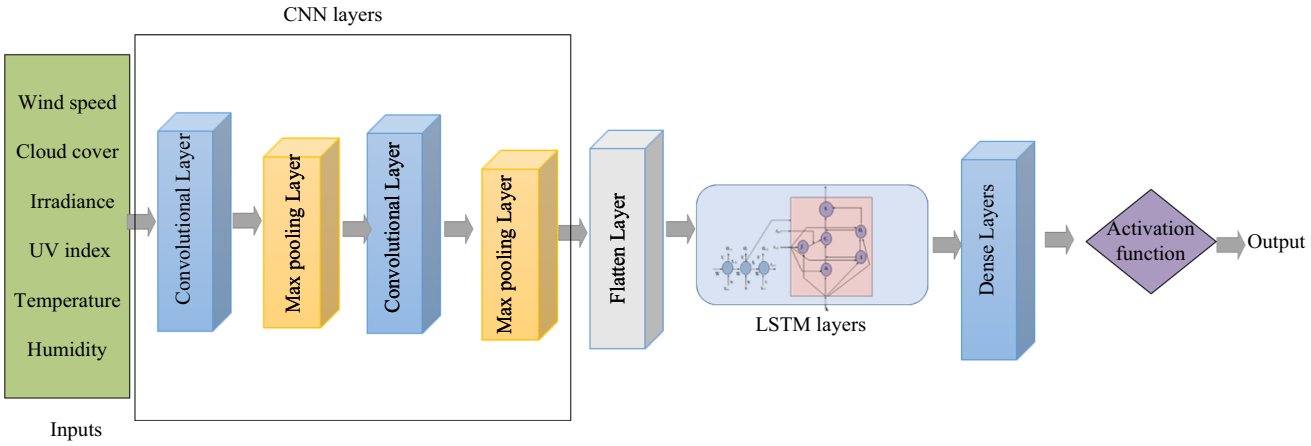


Fig. 4 Structure of CNN-LSTM hybrid model

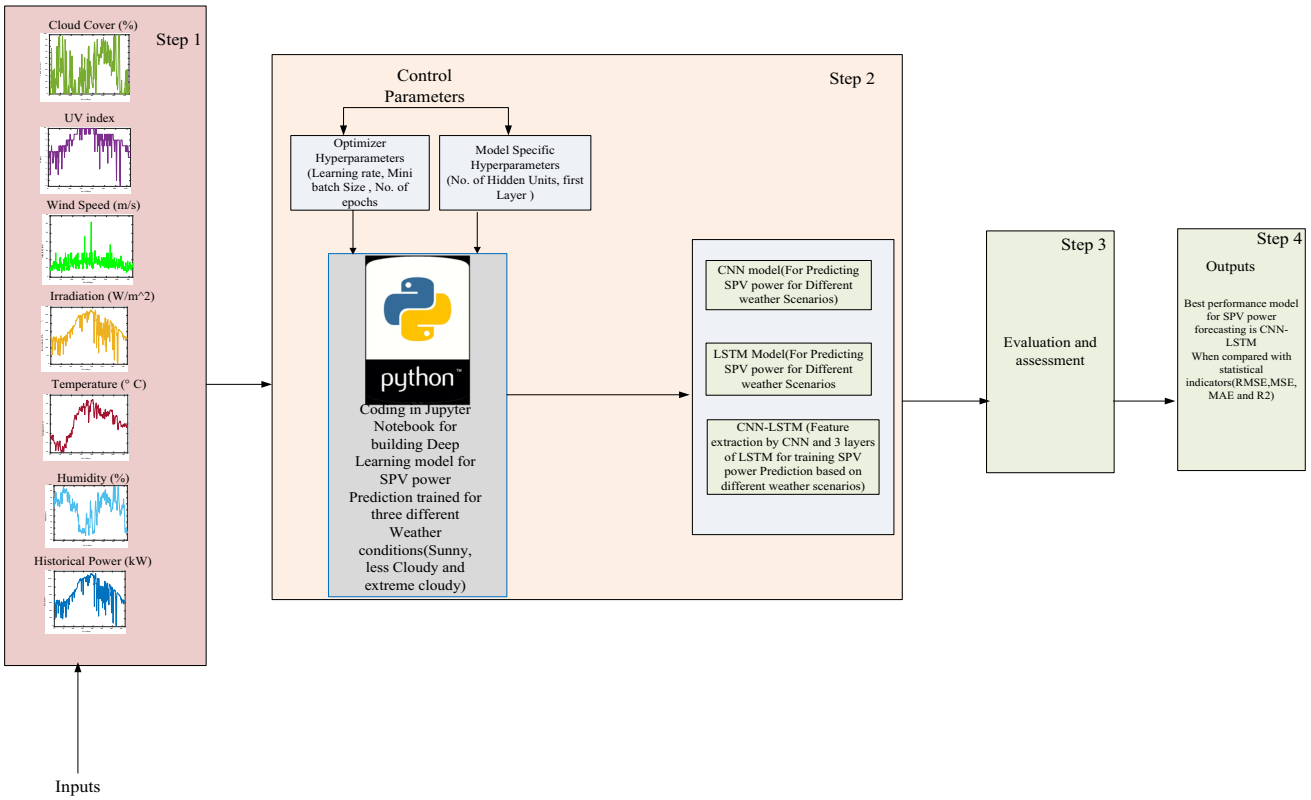


Fig. 5 Flowchart of Proposed work

correlation matrix, the inputs which are having highest correlation with solar PV power are considered as an input for prediction of solar PV power. The multi-inputs chosen in this paper are humidity, temperature, UV index, irradiance, wind speed, cloud cover, and solar PV power which are provided to the CNN, LSTM, and CNN-LSTM model for prediction of solar PV power.

3.1 Selection of inputs-based co-relation matrix

This section explains the inputs used to construct the solar PV power prediction model. The choice of input variables and location is critical to the operation of the PV system.

Despite the flat terrain, the movement of clouds and the kind of clearness index in the sky have an influence on the power output of the PV system. Solar PV power is affected by changes in ambient temperature, humidity, wind speed, cloud cover, UV index, and irradiance. Mathematical computation

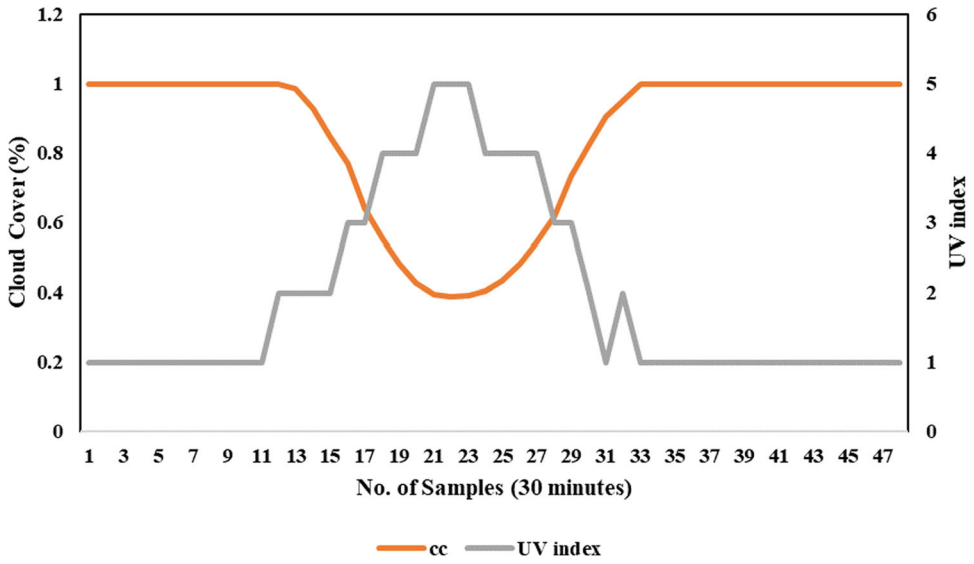


Fig. 6 Cloud cover and UV index relation for 24 h

of cloud cover and UV index is done in this section. The modelling of cloud cover and UV index was done with the help of meteorological parameters such as wind speed, irradiance, humidity, and temperature; these data are recorded on intra-hourly basis so as to get the cloud cover and UV index at 10 min of intervals. The usage of proper sorts of input parameters results in better performing prediction models. The inputs considered for prediction are shown in Fig. 7. This figure is plotted considering one year of data.

If the inputs used are redundant, have a low correlation factor, or essential parameters are missing, the prediction results will be complicated [42].

(A) Modelling of solar PV power

To forecast solar PV power, the real irradiance incident on the PV panel, the exact wind speed on the PV panel, and the PV panel temperature must be estimated. The solar power equation is as follows [16]:

$$P_{\mathrm{c}} = P_{\mathrm{r}} S_m \Big(1 + K_1 \log(S_m) + K_2 \log(S_m)^2 + K_3 T_{\mathrm{a}} + K_4 T_{\mathrm{a}} \log(S_m) + K_5 T_{\mathrm{a}} \log(S_m)^2 + K_6 {T_{\mathrm{a}}}^2 \Big)) \tag{7}$$

$$T_{\mathrm{a}} = \frac{S_m}{26.9 + 6.2 * w_{\mathrm{in}}} \tag{8}$$

$$S_m = \frac{S_{m1}}{1000} \tag{9}$$

$$T_a = T_{a1} - 25^o C \tag{10}$$

$$W_{\mathrm{in}} = \frac{{D_m}^{0.2}}{D_a} * w_s \tag{11}$$

Table 3 Parameters for calculating solar PV power

Parameters	Values
K_1	−0.046689
K_2	−0.072844
K_3	−0.002262
K_4	0.000276
K_5	0.000159
K_6	−0.000006
D_{m}	10 m
D_{a}	7.2 m

where P_c is the real time module estimated power, w_{in} is the speed of wind on the panel, T_{a} is the change in temperature caused by wind speed, and K_1–K_6 are PV module constants. D_{a} is the distance between the PV panel and the ground surface, and D_{m} is the distance between the anemometer and the ground surface. Table 3 shows the values of the constants used to calculate solar PV power. The power received for one year is displayed in Fig. 7g) when P_{r} is set to 5 kW.

(B) Modelling of Cloud Cover.

Cloud cover is the percentage of the sky that is obscured by clouds at a given place and time. It is frequently stated as a percentage or fraction range from 0 to 1, with 0 representing clear sky with no clouds and 1 representing totally overcast sky with clouds covering the whole sky.

Authors in [43] included cloud cover, temperature, relative humidity, and wind speed to develop a model for estimating global solar irradiance. The correlation can be expressed as

$$H = \frac{\left[H_o \sin\alpha_s \left(a_0 + a_1 \frac{N}{10} + a_2 \left(\frac{N}{10}\right)^2 + a_3(T_{\mathrm{a}}) + a_4.RH + a_5 w_s + d\right)\right]}{k} \tag{12}$$

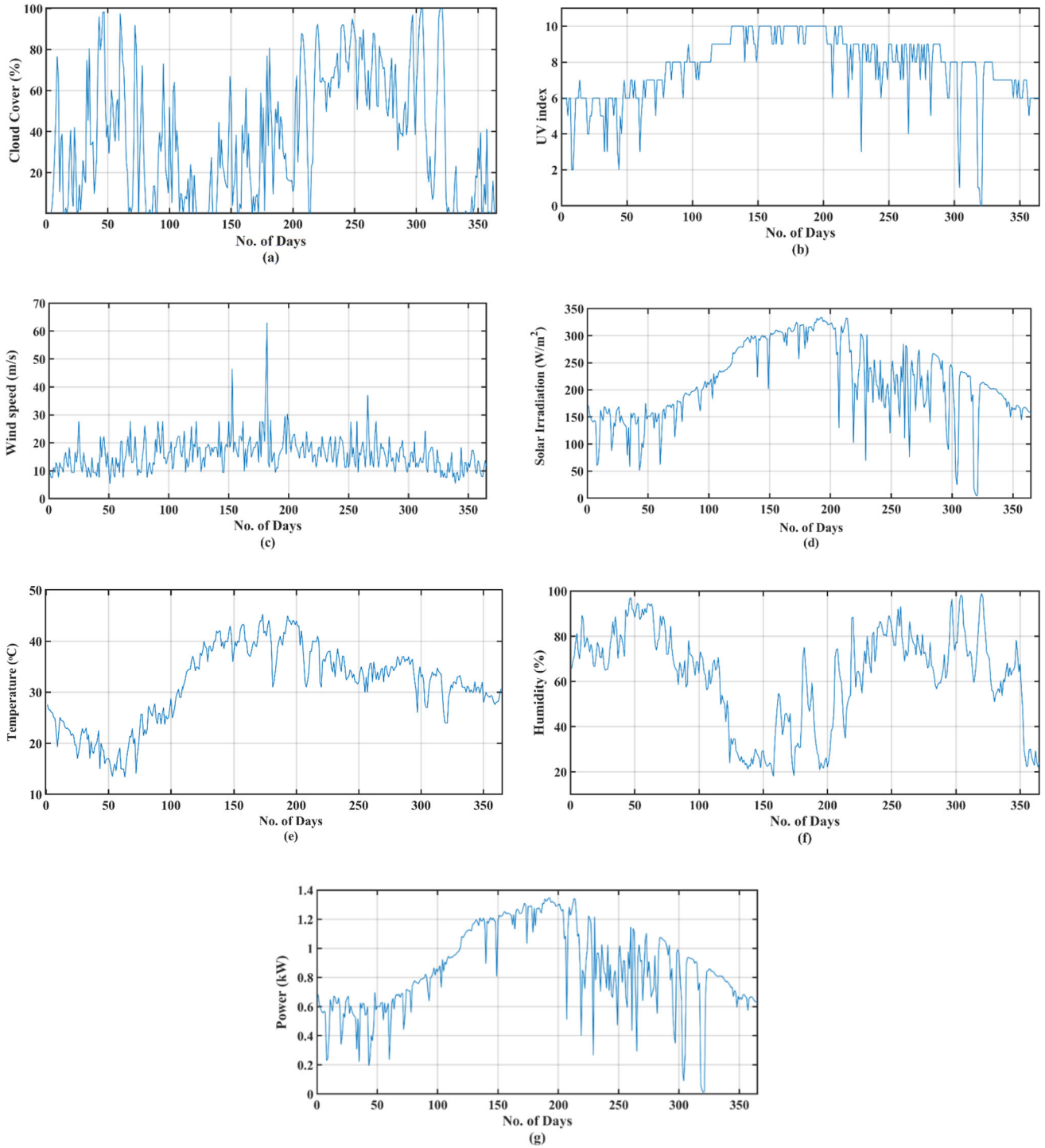


Fig. 7 Daily plots of 1 year for **a** cloud cover (%) **b** UV index **c** wind speed (m/s) **d** Irradiance (W/m.^{2}) **e** Temperature ° C **f** Humidity (%) and **g** power (kW)

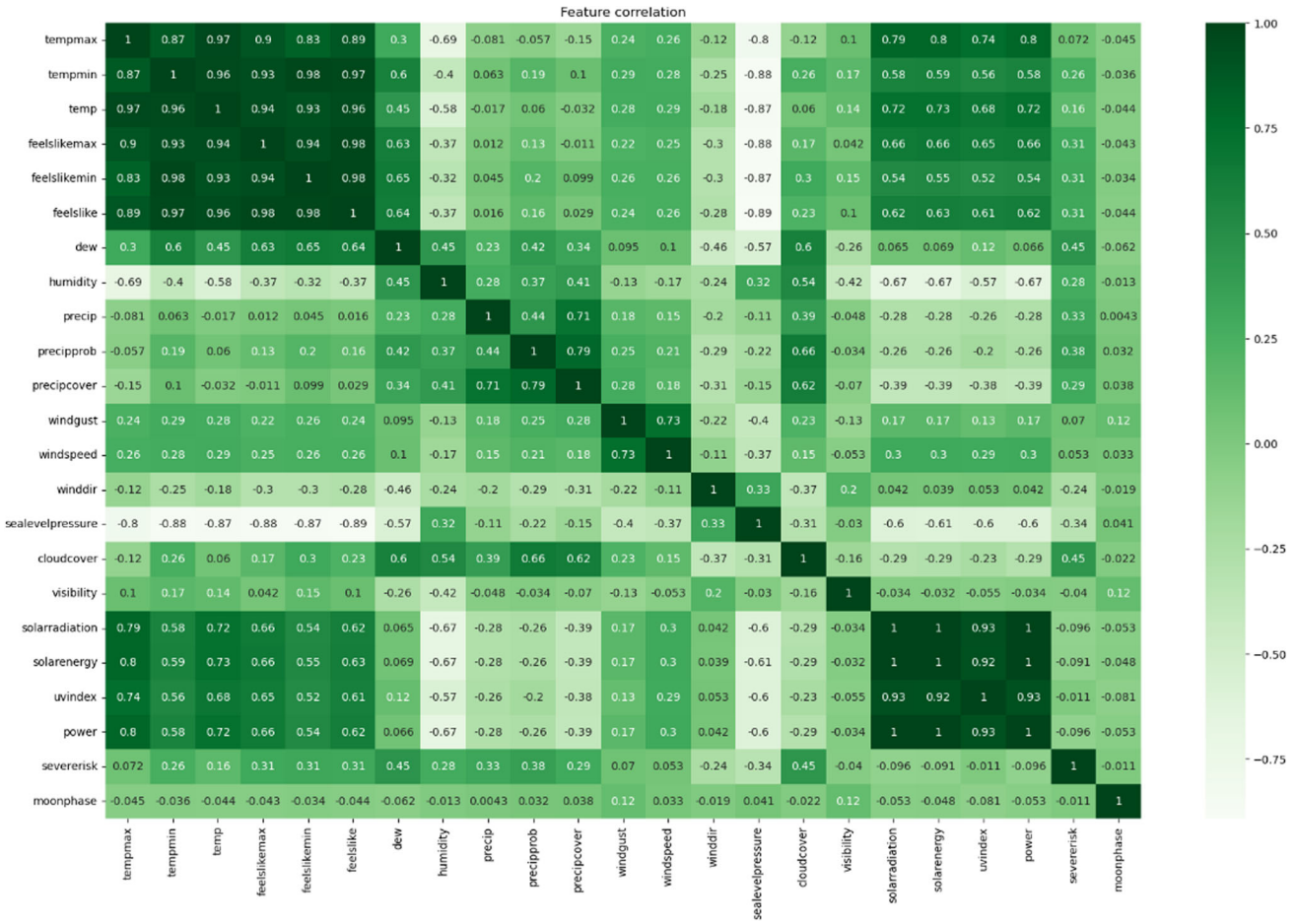


Fig. 8 Feature correlation matrix

There is an inverse relation between cloud cover and irradiance which means more the cloud less will be the irradiance.

where a_0, a_1, a_2, a_3, a_4, a_5, d, and k are regression coefficients; N is the cloud cover in tenths; T_a is the ambient temperature (°C). Here H is hourly global solar radiation (GSR) and H_o is GSR on any horizontal surface under clear sky [44]

There is an inverse relation between cloud cover and Irradiance which means more the percentage of cloud cover less will be the irradiance

$$H = H_o[1 - (1 - k)C] \tag{13}$$

where H_0 represents the expected daily global radiation for a clear sky, C represents the monthly average proportion of the daylight sky covered by clouds, and k represents a constant defining solar radiation transmission inside clouds. This formula has a structure similar to the Angstrom–Prescott (A-P) formula, which describes the relationship between sunlight duration (S) and global solar radiation (H) as:

Calculate H from Eq. (12)

Relation between H and S is as below

$$H = H_o\left[a + b\left(\frac{S}{S_o}\right)\right] \tag{14}$$

a and b are regression coefficients to be found. If the estimate is exact and devoid of aesthetic flaws, the total cloud amount should follow the relationship as [45]:

$$C = \left[1 - \left(\frac{S}{S_o}\right)\right] \tag{15}$$

Nonlinearity was discovered in the connection between C and $(1\text{-}\frac{H}{H_o})$, where H_o represents extraterrestrial solar radiation.

In this work, a relationship between cloud cover and sunlight every 10 min for a one-year New Delhi location data is considered having latitude is 28.749493 longitude is 77.118361. For more accurate cloud cover estimation, sunshine hour data are required.

We can estimate S/S_0 using the correlations reported in (13). Then, multiplying this amount by S_0 gives us the value of S. Cloud cover is calculated from (14).

(C) Modelling of UV index

Table 4 UV index scale range

UV index	Rating
11 or higher	Extreme
8 to 10	Very high
6 to 7	High
3 to 5	Moderate
2 or less	Low

The UV index is scaled from 0 to 11, with 0 being the lowest and 11 being the most intense. These figures are organised as follows:

UV index 2 or less means low, 3 to 5 means moderate, 6 to 7 means high, 8 to 10 means very high, and 11 or higher means extreme as in Table 4; ultraviolet (UV) radiation affects the solar PV power as the correlation with power is 0.8574.

UV relation with solar PV power is linear as the UV is more, power production will be high, and UV index is related to irradiance as in Eq. (16)

$$\mathrm{UVI} = \frac{1}{25}\mathrm{mWm^2UVE} \tag{16}$$

$$\mathrm{UVE} = 40\int_{0.28}^{0.40} E_{\mathrm{gl}}(\lambda)\varepsilon(\lambda)\mathrm{d}\lambda \tag{17}$$

where $E_{gl}(\lambda)$ and $\varepsilon(\lambda)$ are UV global irradiation at the surface and erythemal photobiological global radiation at the surface and the erythemal photobiological reaction action spectrum for human skin at the wavelength (λ). Micrometres (μm) are used to talk about the boundaries of the integral. Each UVI measure is equal to 25 milliwatts per square metre of broadband ultraviolet erythemal energy (UVE). The conversion from UVE (unit in W/m2) to UVI (unitless) makes it easier for people to understand UV levels (WHO—World Health Organisation World Meteorological Organisation United Nations Environment Programme International Commission on Non-Ionizing Radiation Protection, 2002). Since there aren't many UV radiometric standard readings, modelling and satellite remote sensing have been used to determine UVI.

Considering UV index will help in knowing the impact of UV index on solar modules performance. Its correlation with solar PV power is shown in Table 5.

The cloud cover and UV index from Fig. 6 show that when the cloud cover is 1, UV index is minimum, and vice versa, the cloud cover and UV index are modelled with the help of the meterological inputs and the data obtained is at every 30 min.

The correlation of the inputs with solar PV power is shown in Fig. 9 where the irradiance is the best factor which affects the solar PV power as in Table 5. UV index also has high correlation with solar PV power of 0.8574, temperature correlation with solar PV power is 0.6315, and wind speed correlation with power is about 0.08937. Sea-level pressure correlation with solar PV power is 0.3591, humidity and cloud cover show negative correlation which means higher the humidity solar PV power is reduced, and as the percentage of cloud cover increases, the solar PV power is reduced so its impact on solar power for prediction is an important factor. Dataset taken from solcast site has several parameters, and the feature correlation matrix is shown in Fig. 8. Table 5 depicts the relationship between the six specified inputs and the power of the PV system.

Table 5 Correlation coefficient between the six selected inputs to the solar PV power

Inputs	Correlation Coefficient (R^2)
Irradiance	1
Temperature	0.6315
Humidity	−0.6509
Wind Speed	0.3937
Cloud cover	−0.2931
UV index	0.8574

4 Results and discussion

Deep learning models and proposed hybrid models are employed for the prediction of solar power using multivariate time series data. The selection of the inputs that are utilised in the models that are run through deep learning is covered in Sect. 4. In the first place, a hybrid model that consists of CNN and LSTM features is presented. The dataset that was used for this analysis spanned one year containing 52,560 (365*24*6), and additional LSTM layers were connected in order to learn the pattern of solar PV power prediction based on weather scenarios, i.e. sunny, partially cloudy, and extremely cloudy. The CNN model assisted in the extraction of features from the dataset. The outcomes that were predicted are then compared with the LSTM and CNN models. In this section, the dataset was divided into four seasons: winter, summer (pre monsoon), rainy, and post-monsoon. Deep learning models such as CNN, LSTM, CNN-LSTM, and ARIMA are trained to predict solar power based on the seasonally divided dataset. To further improve the accuracy of solar power predictions, the models were trained on data divided by different weather patterns. This section is divided into three parts: The first part discusses the prediction based on seasonal patterns, i.e. (winter, summer, rainy, and post monsoon), and then the second shows the prediction based

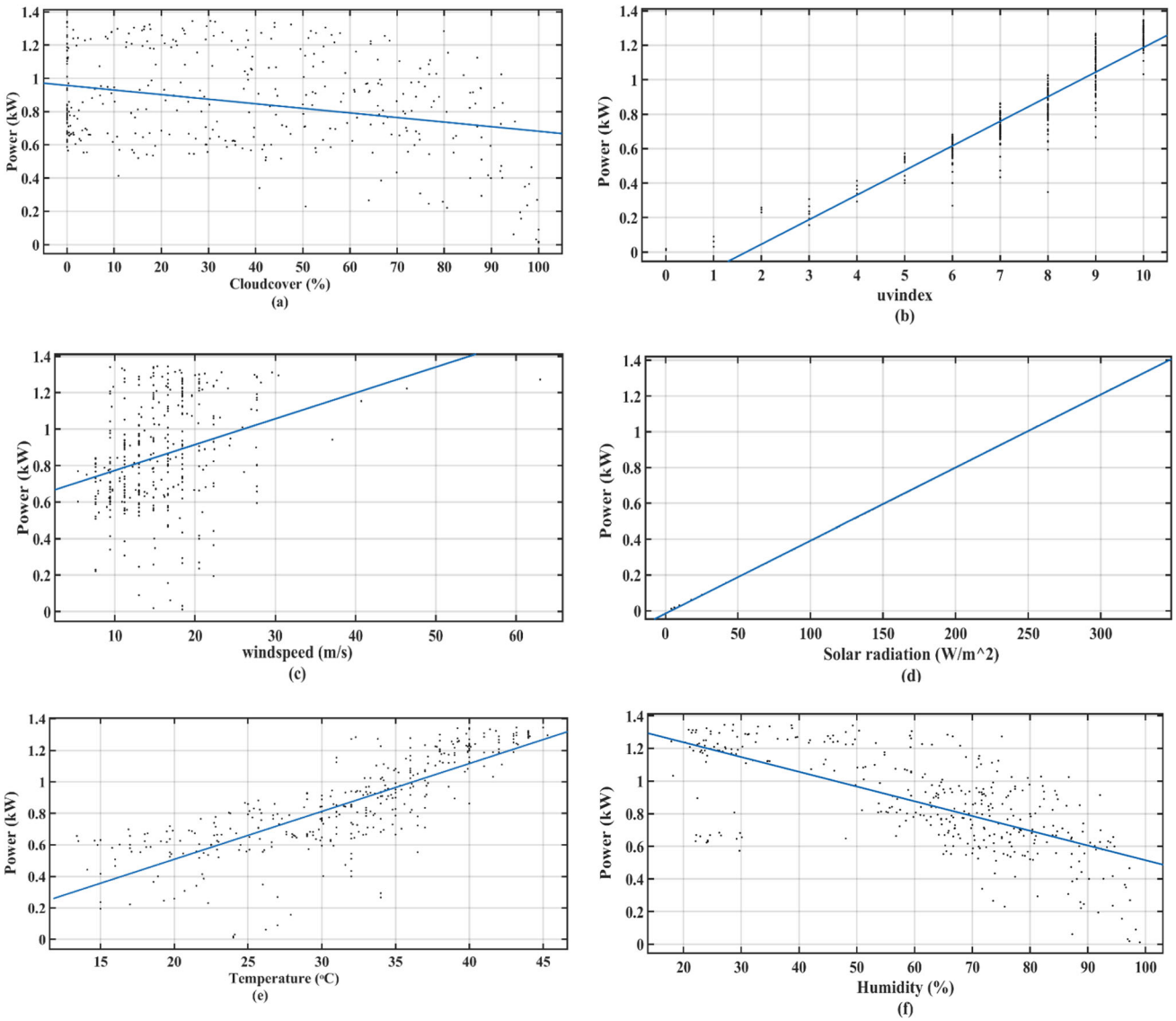


Fig. 9 Correlation plot of solar power with **a** cloud cover **b** UV index **c** wind speed **d** Irradiance **e** Temperature **f** Humidity

on three weather conditions, and the third part discusses the performance metrices.

Data are collected throughout a year and used to train deep learning models such as CNN-LSTM, CNN, and LSTM. In addition, the ARIMA model is trained to predict solar power. The dataset contains 365 data points for a year. Figure 10a depicts a comparison prediction plot of deep learning models. The error plots are shown in Fig. 10b. Based on the data shown in Fig. 10a, it can be concluded that the CNN-LSTM model is a better approach for predicting solar output, with less mistakes than other proposed methods.

4.1 Solar power prediction based on different seasons

The complete dataset is divided in four different seasons—winter, summer (pre monsoon), rainy (monsoon) and post monsoon—

The winter season encompasses the months of December to March.

The summer (premonsoon) season spans from April to June.

The monsoon season, known as the rainy season, spans from June to September.

And the post-monsoon season spans from October to December.

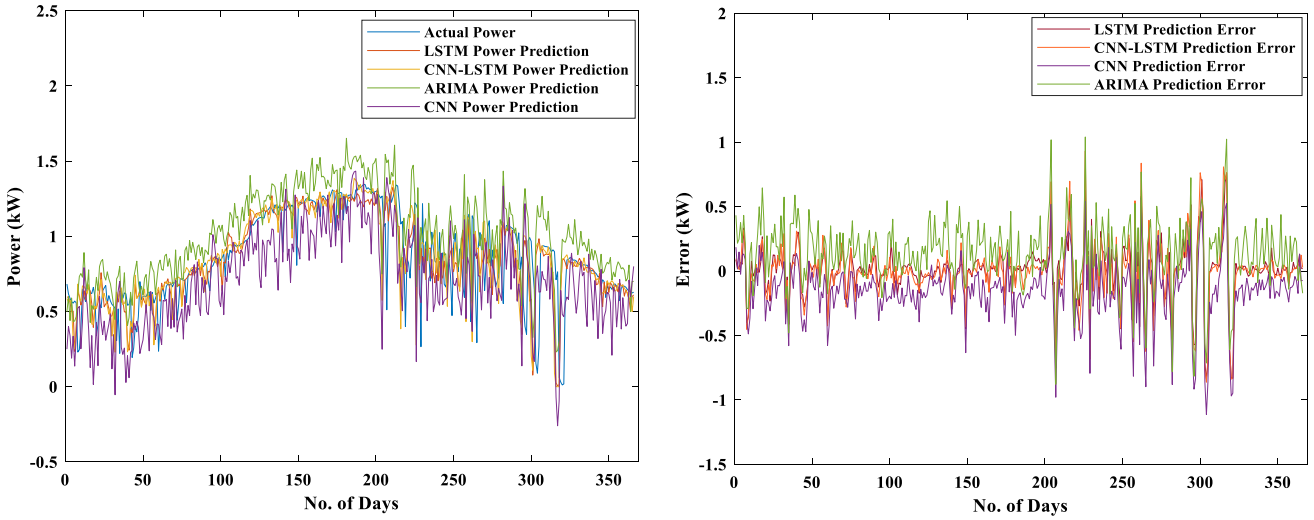


Fig. 10 Comparison of CNN-LSTM, LSTM, CNN and ARIMA models for one year data **a** solar PV power prediction plot with actual power **b** error plots

Table 6 Distributions of different weather conditions

UV index	Type of weather conditions
0% to 20%	Clear day (sunny)
21% to 60%	Partially Cloudy
61% to 100%	Extremely Cloudy

These different seasonal pattern datasets are used to train the deep learning models for different seasons. The deep learning models are trained based on different seasonal pattern and solar power prediction is made according to different seasons. From Fig. 11 it can be observed that maximum power is obtained in months of summer the variations are also less as compared to rainy, post-monsoon and winter conditions. Figure 11 shows the solar power prediction comparison plot and error plot of various deep learning models with the traditional model i.e. ARIMA.

4.2 Solar power prediction based on weather conditions

One year of solar power data is analysed based on three different weather conditions; these weather conditions are obtained based on Table 6.

Based on Table 6 the dataset is divided into 3 different weather scenarios. The sunny dataset has 144 data points, the partially cloudy dataset has 141 days, and the extremely cloudy dataset has 81 days.

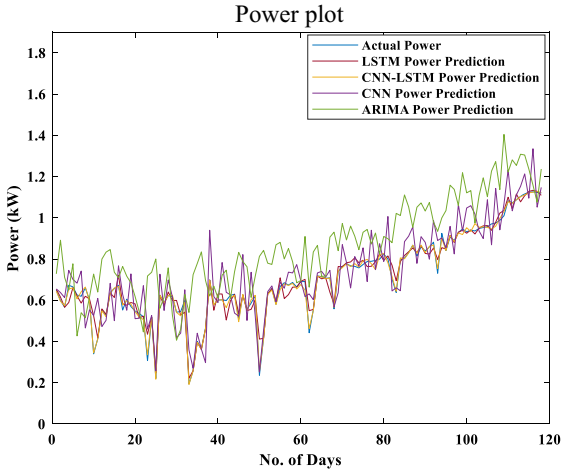
4.2.1 Solar PV power prediction for sunny day

The data were collected from solcast site for one year where sunny condition data were used to train the three models, i.e. CNN, LSTM and the proposed hybrid CNN-LSTM. The sunny data are 144; these data are used to train the deep learning models for solar PV power prediction. From Fig. 12a it can be observed when there is sun in sky, power increases or decreases gradually in the sky. Seven inputs were given to all the three model for analysing; they were humidity, temperature, wind speed, solar radiation, UV index cloud cover, and solar PV power; the predicted output is solar PV power; the forecasted results for all the three models are shown in Fig. 12a; and it is observed that CNN-LSTM predicted more accurately than CNN model and LSTM model. The error between all the three models has been compared which shows CNN-LSTM has less errors than the other two models, the error plot is shown in Fig. 12b, and it is observed that CNN-LSTM have less error than other CNN and LSTM model.

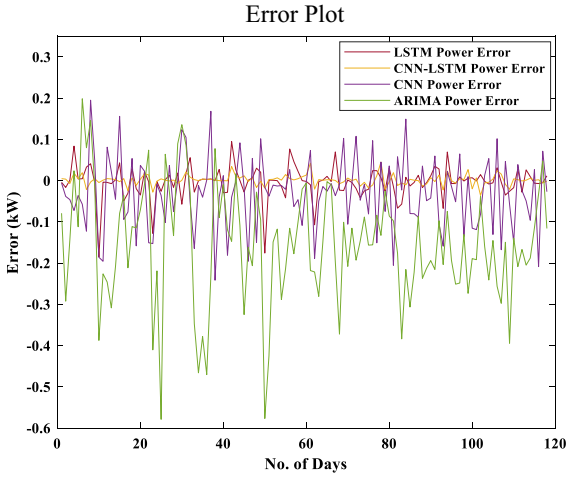
Traditional methods of prediction are compared for sunny conditions in order to show that the proposed deep learning model performs much better than traditional model; the traditional model considered is ARIMA (auto-regressive integrated average). For comparison MSE is calculated for ARIMA model, it comes out to be 0.00220, and MSE for proposed CNN-LSTM model comes out to be 0.000645. The comparison is shown in Fig. 12a.

Additionally, the models are compared using root mean square error, mean absolute error, mean square error, and R^2 for the goal of conducting more accurate analyses.

Winter season

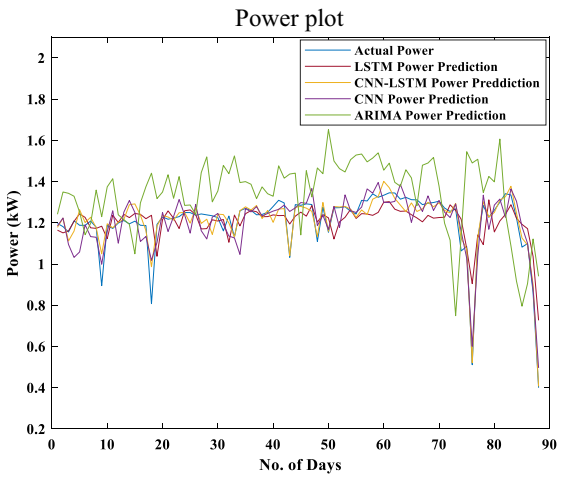


(a)

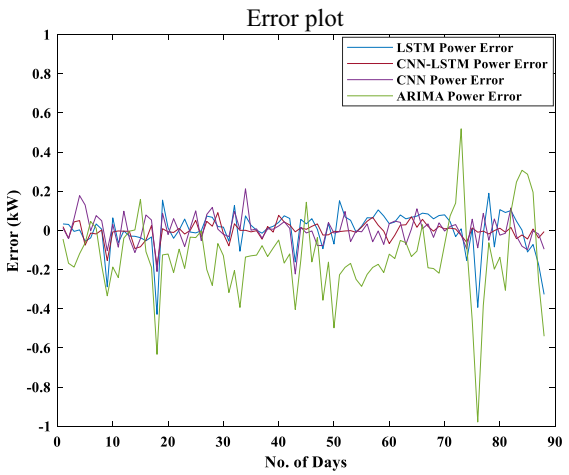


(b)

Summer season

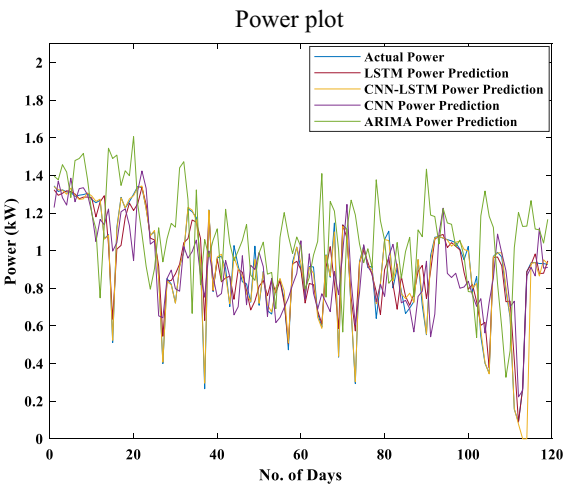


(c)

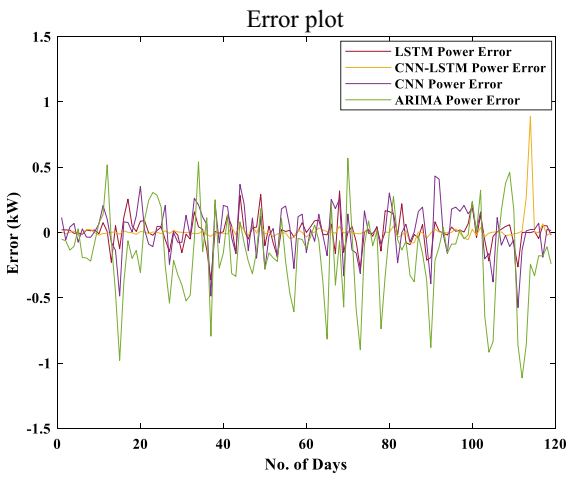


(d)

Rainy season



(e)



(f)

Fig. 11 Comparison of different deep learning Models with ARIMA model in four different seasons

Post Monsoon Season

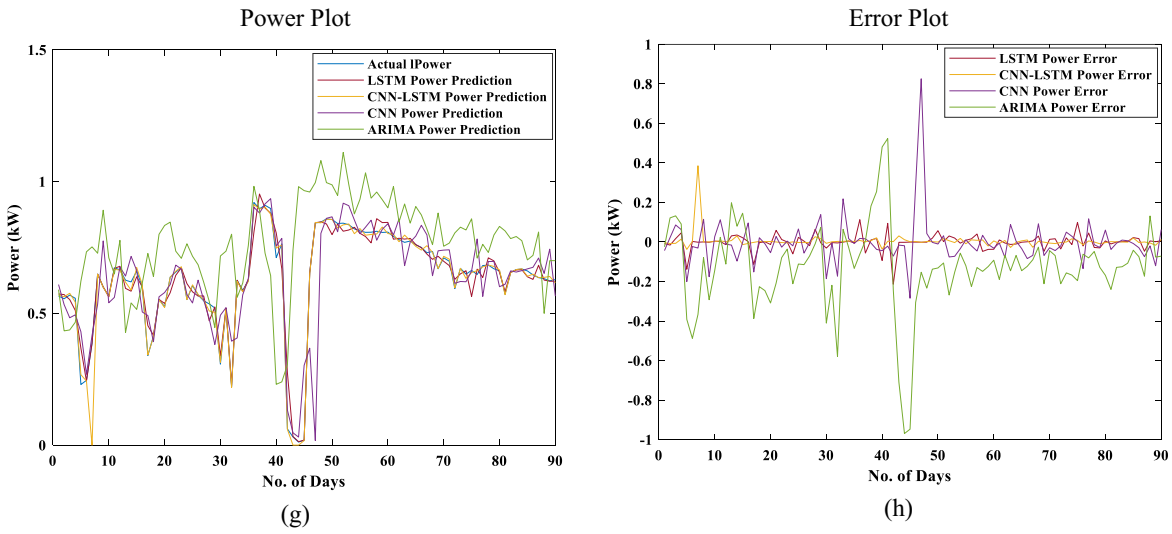


Fig. 11 continued

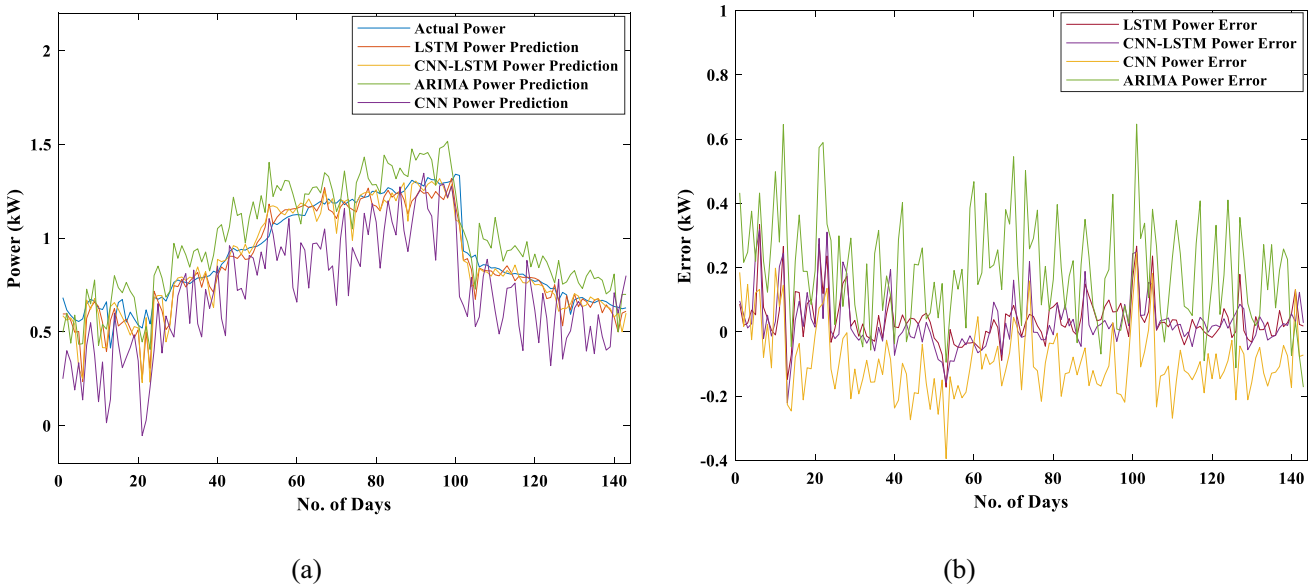
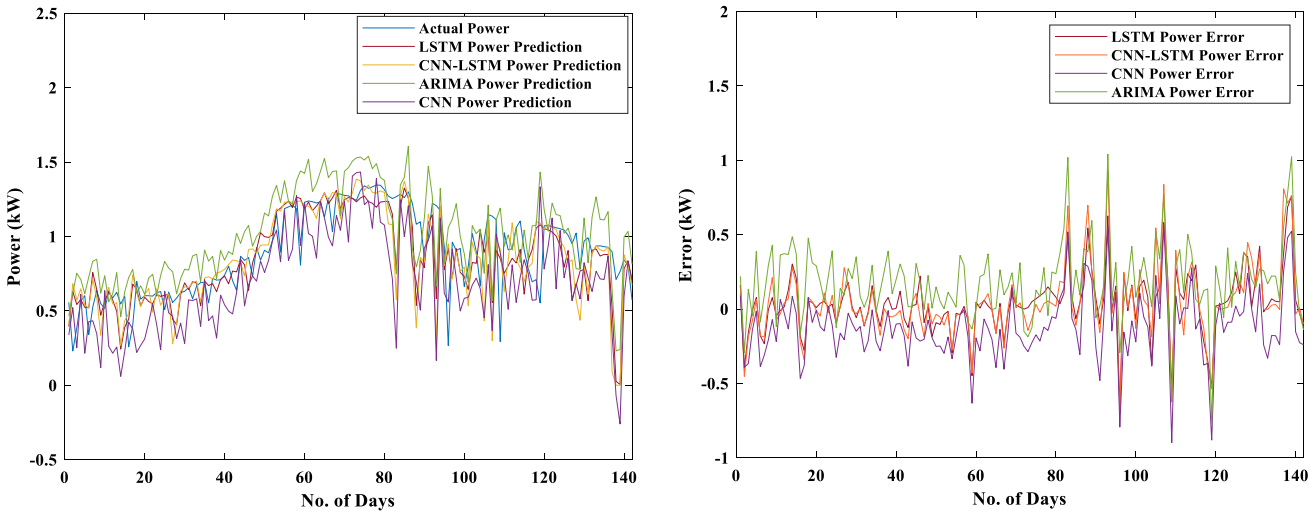


Fig. 12 Comparison of CNN-LSTM, LSTM, CNN, and ARIMA models in sunny conditions **a** solar PV power prediction plot with actual power **b** error plots

4.2.2 Solar PV power prediction with partially cloudy conditions

The data were collected from solcast site for one years where partially cloudy condition data were used to train the three models, i.e. CNN, LSTM and the proposed hybrid CNN-LSTM. The partially cloudy data are 141; these data are used to train the deep learning models for solar PV power prediction. From Fig. 13a it can be observed when there are clouds in the sky, there is an immediate shift in the amount of fluctuation in the power generation. In cloudy conditions the solar PV power fluctuated more as compared to Sunny conditions. Seven inputs were given to all the three models for analysing; they were humidity, temperature, wind speed, solar radiation, UV index cloud cover, solar PV power and the predicted output solar PV power; the forecasted results for all the three models are shown in Fig. 13. It is observed that CNN-LSTM predicted more accurately than CNN model and LSTM model. The error between all the three models has been compared which shows CNN-LSTM has less errors than the other two models and the error plot is shown in Fig. 13b. It



(a) (b)

Fig. 13 Comparison of CNN-LSTM, LSTM, CNN, and ARIMA model in partially cloudy conditions **a** Solar PV power prediction plot with actual power **b** error plots

can be observed that CNN-LSTM error is very less compared to CNN and LSTM model.

4.2.3 Solar PV power prediction with extremely cloudy conditions

The data were collected from solcast site for one years where extremely cloudy condition data were used to train the three models, i.e. CNN, LSTM, and the proposed hybrid CNN-LSTM. The extremely cloudy data are 81; these data are used to train the deep learning models for solar PV power prediction. In extremely cloudy condition, frequent fluctuation of solar PV power is observed. And the solar PV power generated is less in extremely cloudy conditions. From Fig. 14a it can be observed when there are clouds in the sky, there is an immediate shift in the amount of fluctuation in the power generation. Seven inputs were given to all the three models for analysing; they were humidity, temperature, wind speed, solar radiation, UV index cloud cover, solar PV power, and the predicted output solar PV power. The forecasted results for all the three models are shown in Fig. 14; it is observed that CNN-LSTM predicted more accurately than CNN model and LSTM model. The error between all the three models has been compared which shows CNN-LSTM has less errors than the other two models; the error plot is shown in Fig. 14b. It can be observed that CNN-LSTM error is very less compared to CNN and LSTM model. Traditional methods of prediction are compared for extremely cloudy conditions in order to show that the proposed deep learning model performs much better than traditional model; the traditional model considered is ARIMA. For comparison MSE is calculated for ARIMA model, it comes out to be 0.00820, and MSE for proposed CNN-LSTM model comes out to be 0.0012. The comparison is shown in Fig. 14.

4.3 Performance Evaluation

Performance estimation is an essential factor for determining a model's prediction accuracy. Root means square error (RMSE), moving average error (MAE), mean square error (MSE), and R^2 are variables used for comparison analysis. MAE aids in analysing the prediction model's constant prediction errors, RMSE assesses the model's overall accuracy, and R^2 represents the relationship strength between the predicted and actual values of the models. Table 7 shows partially cloudy, extremely cloudy, and sunny prediction error.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|A_i - P_i| \tag{18}$$

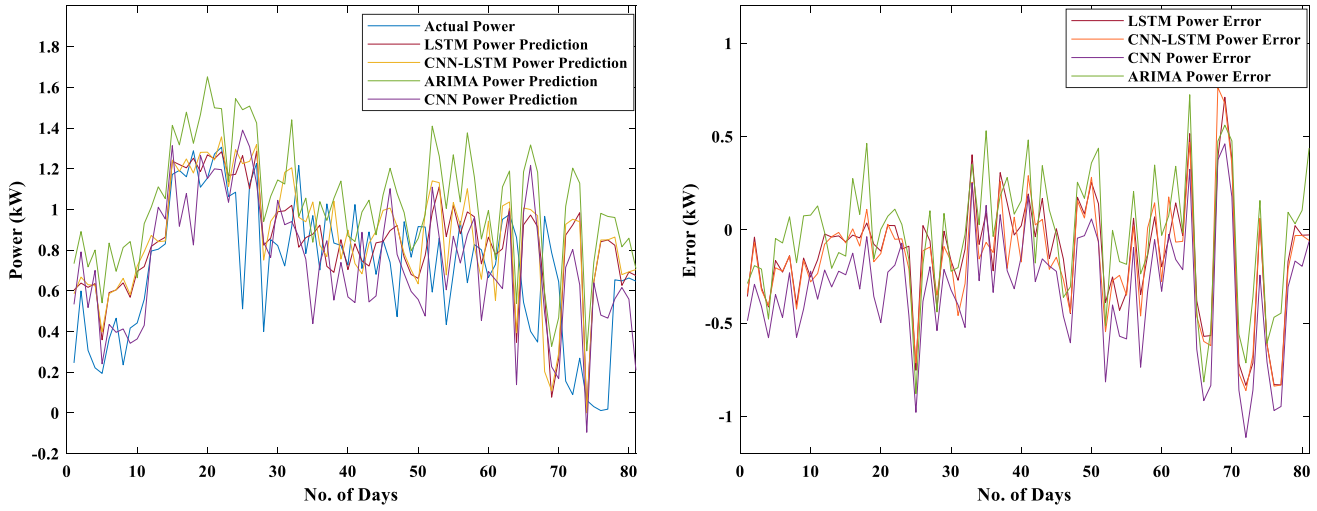
$$\mathrm{NMAE} = \frac{\mathrm{MAE}}{\mathrm{Range}(A_i)} \tag{19}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(A_i - P_i)^2 \tag{20}$$

$$\mathrm{NMSE} = \frac{\mathrm{MSE}}{\mathrm{Range}(A_i)} \tag{21}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - Pi)^2} \tag{22}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\mathrm{Range}(A_i)} \tag{23}$$



(a) (b)

Fig. 14 Comparison of CNN-LSTM, LSTM, CNN, and ARIMA models in extremely cloudy conditions **a** solar PV power prediction plot with actual Power **b** error plots

Table 7 Estimation errors of various prediction model with weather conditions such as sunny, partially cloudy and extremely cloudy

Model (Sunny)	MAE	NMAE	RMSE	NRMSE	MSE	NMSE	R^2
LSTM	0.0286	0.0308	0.04800	0.05175	0.00230	0.00248	0.9636
CNN	0.0271	0.0292	0.04406	0.04750	0.00194	0.00209	0.9643
CNN-LSTM	0.0163	0.0176	0.02541	0.02739	0.00064	0.00069	0.9898
ARIMA	0.435	0.469	0.05400	0.05821	0.00220	0.00237	0.9239
Model (Partially Cloudy)	MAE	NMAE	RMSE	NRMSE	MSE	NMSE	R^2
LSTM	0.9951	0.7420	0.14622	0.1090	0.02138	0.01594	0.7734
CNN	0.0919	0.0685	0.1237	0.0922	0.1529	0.1140	0.8378
CNN-LSTM	0.0236	0.0175	0.0346	0.0258	0.00120	0.00089	0.9872
ARIMA	0.1234	0.0920	0.19622	0.1463	0.00820	0.00611	0.7234
Model (Extremely Cloudy)	MAE	NMAE	RMSE	NRMSE	MSE	NMSE	R^2
LSTM	0.0719	0.0555	0.1046	0.0808	0.0109	0.0084	0.8967
CNN	0.0535	0.0413	0.1246	0.0962	0.03186	0.02461	0.7045
CNN-LSTM	0.2521	0.1947	0.0824	0.06366	0.00679	0.00524	0.9358
ARIMA	0.1576	0.1218	0.2146	0.1658	0.0423	0.0327	0.6045

$$R^2 = 1 - \frac{\sum_k (A_i - P_i)^2}{\sum_k \left(A_i - \overline{A}\right)^2} \tag{24}$$

where "N" number of windows, "A_i" actual value, and P_i does the Predicted value. The arithmetic mean of the actual value is $\overline{A}$.

Tables 7 and 8 analyse the performance of proposed methodology for varying weather conditions and seasons, respectively. According to Table 7, the CNN-LSTM algorithm's performance is superior than that of both CNN and LSTM which performs better than LSTM in the majority of deep learning applications; nevertheless, when compared with the CNN-LSTM model, its performance is less accurate in terms of performance evaluation. It is possible to draw the conclusion from Table 7 that the deep learning model is superior to cloudy weather circumstances in terms of its ability to forecast. Furthermore, it is possible to deduce from Table 8 that the models have a high level of accuracy when applied to the winter and post-monsoon datasets. In Tables 7 and 8, deep learning models perform better than the ARIMA.

CNN-LSTM is the deep learning model that performs the best across all of the weather patterns and seasons that were taken into consideration. In Table 9, the work of previous

Table 8 Estimation errors for various prediction models trained based on different seasons

Model (Winter)	MAE	NMAE	RMSE	NRMSE	MSE	NMSE	R^2
LSTM	0.02424	0.02601	0.0395	0.04237	0.00155	0.00166	0.9628
CNN	0.06323	0.06783	0.08511	0.09131	0.00724	0.00776	0.8271
CNN-LSTM	0.00862	0.00924	0.0126	0.0135	0.000158	0.000169	0.9962
ARIMA	0.0525	0.05632	0.08900	0.09548	0.00780	0.008368	0.9239
Model (Summer)	MAE	NMAE	RMSE	NRMSE	MSE	NMSE	R^2
LSTM	0.06974	0.07376	0.1037	0.10968	0.01075	0.01137	0.7225
CNN	0.05405	0.05716	0.0723	0.0765	0.00524	0.00554	0.7674
CNN-LSTM	0.02869	0.03034	0.0425	0.0449	0.00181	0.00191	0.9197
ARIMA	0.23363	0.24710	0.2789	0.2949	0.0123	0.01301	0.5674
Model (Rainy)	MAE	NMAE	RMSE	NRMSE	MSE	NMSE	R^2
LSTM	0.07394	0.05893	0.1094	0.08719	0.01196	0.00953	0.8396
CNN	0.14043	0.11193	0.18216	0.14519	0.03318	0.02645	0.6554
CNN-LSTM	0.02592	0.02066	0.08858	0.07060	0.00784	0.00625	0.8947
ARIMA	0.15578	0.1242	0.2345	0.1869	0.004323	0.00344	0.5556
Model (Post Monsoon)	MAE	NMAE	RMSE	NRMSE	MSE	NMSE	R^2
LSTM	0.02599	0.03228	0.0429	0.05298	0.0018	0.002236	0.95078
CNN	0.06499	0.08074	0.1195	0.1484	0.01429	0.01775	0.71822
CNN-LSTM	0.01203	0.0149	0.04225	0.0525	0.00178	0.00221	0.9523
ARIMA	0.07157	0.0889	0.24146	0.2998	0.01523	0.0189	0.6238

author is compared with the proposed, and the results show that the proposed model is more effective.

5 Conclusion

A prediction model for PV power generation can enhance the design, operation, and stability of PV power plants by making more accurate forecasts based on meteorological data. But because the weather changes all the time, it is difficult to know how much PV power is produced. To make a good prediction model that only uses past data and leaves out data on solar radiation that is highly correlated with PV power production, you need a statistical method for figuring out how past and short-term data depend on each other. In this study, a CNN-LSTM hybrid model is proposed for estimating how much PV power is produced. The suggested model addresses the shortcomings of the previous models while retaining their benefits. The proposed model has been compared to other deep learning models in which only LSTM and CNN models are given multiple inputs. However, their performance is less accurate than the hybrid model because in the proposed hybrid model, the CNN model first learns the pattern of different weather conditions and then separates LSTMs for prediction solar PV power which help the model make more accurate prediction of PV power. A quantitative evaluation also showed that the proposed model has RMSE of 0.0254, 0.03465, and 0.0824, MAE of 0.0163, 0.0236, and 0.2521, MSE of 0.000645, 0.00120, and 0.00679, and R^2 of 0.9898, 0.9872, and 0.9358 for sunny, partially cloudy, and extremely cloudy day data, respectively. The normalised RMSE, MAE, and MSE came out to be 0.0176, 0.02739, and 0.00069 for sunny conditions. For partially cloudy conditions the normalised RMSE, MAE, and MSE came out to be 0.0175, 0.0258, and 0.00089 and for extremely cloudy conditions the normalised RMSE, MAE, and MSE came out to be 0.1947, 0.06366, and 0.00524. This model works better when the weather is sunny. When different season datasets are considered, the proposed CNN-LSTM models perform better in winter seasons and worst in rainy season. It is seen that PV power is increased in sunny conditions and decreased in non-ideal conditions, i.e. partially cloudy and extremely cloudy. The amount of power a PV power plant can make depends on the size of the installation and the weather, so the model needs to be changed. In a future study, the power-generating capacity of PV power plants should be changed to predict solar PV power. To make forecasts more accurate, optimisation methods automatically change models based on data from the system. Solar radiation and the properties of PV panels must also be analysed.

Table 9 Comparison of the proposed CNN-LSTM method with previous work

Ref	Method Proposed	Input Parameters	Dataset	Performance Metric	Observation
[46]	CNN-LSTM with a Semi Asynchronous Personalised Federated Learning Framework	Not given	990 days	0.689 (RMSE)	The suggested system greatly enhanced performance without exchanging raw data and further improved the PV power production performance projection
[47]	CNN-LSTM	Time sampling	1 month	0.03105 (RMSE)	CNN LSTM outperforms LSTM method
[48]	Deep ESN and CNN-Dee ESN	Wind speed, humidity, output power, global horizontal radiation, diffuse radiation, average phase current, and temperature	11 months	0.04101(RMSE) 0.0381(MAE) 3.3313 (MAPE)	In solar power generation prediction, the suggested hybrid model beat other single models and attained the lowest performance measure
[19]	CNN-LSTM	Solar power	1 year	0.485 and 0.2775 RMSE for sunny and cloudy, respectively	CNN-LSTM Outperforms in sunny Weather conditions
[49]	CNN-LSTM	Humidity Wind speed, output power, diffuse radiation, global horizontal radiation, average phase current, and temperature	–	O.623 (RMSE)	The suggested hybrid prediction model outperforms the single model LSTM, CNN
[50]	MODWT-LSTM	Wind speed, wind direction, Active power, temperature, humidity, global horizontal radiation, diffuse horizontal radiation, and daily rainfall	1 year	0.1231 (RMSE)	As compared to 1 day and 10 days, the suggested technique is found to be more dependable and to have the best produced efficient accuracy over a long-term prediction horizon of 1 month
	Proposed model	Temperature, humidity, wind speed, cloud cover, UV index, solar irradiance, Solar power	12 Months	0.025407 RMSE (sunny) 0.03465 RMSE (partially cloudy) 0.0824 RMSE (Extremely cloudy)	

Acknowledgements The authors would like to thank Science and Engineering Research Board (SERB) to carry out this research under CRG scheme (CRG/2022/003902).

Author contributions Rahma Aman is the corresponding author, and M. Rizwan and Astitva Kumar reviewed the manuscript

Data availability Dataset analysis and prediction have been performed utilizing publicly available data that have been gathered from solcast site for one year from November 2021 to November 2022.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval The authors declare that they have no financial or personal connections to any individual or entity that may have an undue influence on this work. This document is the author's original work and does not contain any material borrowed from another source. All data measurements are genuine, unmodified results, and this work has never been published or submitted to another journal for publication consideration. Furthermore, none of the authors have any financial or scientific conflicts of interest in relation to the study presented in this publication.

References

1. Vyas S,. Goyal Y, Bhatt N, Bhuwania S, Patel H, Mishra S, Tripathi B (2022) Prediction solar power generation on the basis of predictive and corrective maintenance activities https://doi.org/10.48550/arxiv.2205.08109
2. Sharma B, Rizwan M, Anand P (2023) Optimal design of renewable energy based hybrid system considering weather forecasting using machine learning techniques. Electr Eng. https://doi.org/10.1007/s00202-023-01945-w

3. Huang X, Liu J, Xu S, Li C, Li Q, Tai Y (2023) A 3D ConvLSTM-CNN network based on multi-channel color extraction for ultra-short-term solar irradiance forecasting. Energy 272:127140. https://doi.org/10.1016/J.ENERGY.2023.127140
4. Neshat M, Nezhad MM, Mirjalili S, Garcia DA, Dahlquist E, Gandomi AH (2023) Short-term solar radiation forecasting using hybrid deep residual learning and gated LSTM recurrent network with differential covariance matrix adaptation evolution strategy. Energy 278:127701. https://doi.org/10.1016/j.energy.2023.127701
5. van der Meer DW, Widén J, Munkhammar J (2018) Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 81:1484–1512. https://doi.org/10.1016/J.RSER.2017.05.212
6. Hussain T, Min Ullah FU, Muhammad K, Rho S, Ullah A, Hwang E, Moon J, Wook Baik S (2021) Smart and intelligent energy monitoring systems: a comprehensive literature survey and future research guidelines. Wiley Online Libr. 45:3590–3614. https://doi.org/10.1002/er.6093
7. Li H, Wang X, Zhang K, Wu S, Xu Y, Liu Y, Qiu C, Zhang J, Fu E, Li L (2021) A neural network-based approach for the detection of heavy precipitation using GNSS observations and surface meteorological data. J Atmos Solar-Terrestrial Phys 225:105763. https://doi.org/10.1016/j.jastp.2021.105763
8. Ziane A, Necaibia A, Sahouane N, Dabou R, Mostefaoui M, Bouraiou A, Khelifi S, Rouabhia A, Blal M (2021) Photovoltaic output power performance assessment and forecasting: Impact of meteorological variables. Sol Energy 220:745–757. https://doi.org/10.1016/j.solener.2021.04.004
9. De Giorgi MG, Congedo PM, Malvoni M (2014) Photovoltaic power forecasting using statistical methods: impact of weather data. IET Sci Meas Technol 8:90–97. https://doi.org/10.1049/IET-SMT.2013.0135
10. Wu YK, Huang CL, Phan QT, Li YY (2022) Completed review of various solar power forecasting techniques considering different viewpoints. Energies 15:3320. https://doi.org/10.3390/EN15093320
11. Mellit A, Massi Pavan A, Lughi V (2014) Short-term forecasting of power production in a large-scale photovoltaic plant. Sol Energy 105:401–413. https://doi.org/10.1016/j.solener.2014.03.018
12. Ait Maatallah O, Achuthan A, Janoyan K, Marzocca P (2015) Recursive wind speed forecasting based on hammerstein autoregressive model. Appl Energy 145:191–197. https://doi.org/10.1016/J.APENERGY.2015.02.032
13. Kómar L, Kocifaj M (2016) Statistical cloud coverage as determined from sunshine duration: a model applicable in daylighting and solar energy forecasting. J Atmos Solar-Terrestrial Phys 150–151:1–8. https://doi.org/10.1016/j.jastp.2016.10.011
14. Deo RC, Wen X, Qi F (2016) A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl Energy 168:568–593. https://doi.org/10.1016/J.APENERGY.2016.01.130
15. Mandal P, Madhira STS, Ulhaque A, Meng J, Pineda RL (2012) Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. Procedia Comput. Sci. 12:332–337. https://doi.org/10.1016/J.PROCS.2012.09.080
16. Obiora CN, Hasan AN, Ali A, Alajarmeh N (2021) Forecasting hourly solar radiation using artificial intelligence techniques. IEEE Can J Electr Comput Eng 44:497–508. https://doi.org/10.1109/ICJECE.2021.3093369
17. Abuella M, Chowdhury B (2017) Random forest ensemble of support vector regression models for solar power forecasting, In: 2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference ISGT 2017. https://doi.org/10.1109/ISGT.2017.8086027.
18. Benali L, Notton G, Fouilloy A, Voyant C, Dizene R (2019) Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components. Renew Energy 132:871–884. https://doi.org/10.1016/J.RENENE.2018.08.044
19. Lim SC, Huh JH, Hong SH, Park CY, Kim JC (2022) Solar power forecasting using CNN-LSTM hybrid model. Energies 15:8233. https://doi.org/10.3390/EN15218233
20. Liu X, Liu Y, Kong X, Ma L, Besheer AH, Lee KY (2023) Deep neural network for forecasting of photovoltaic power based on wavelet packet decomposition with similar day analysis. Energy 271:126963. https://doi.org/10.1016/J.ENERGY.2023.126963
21. Tovar M, Robles M, Rashid F (2020) PV power prediction, using CNN-LSTM hybrid neural network model. Case of Study: Temixco-Morelos, México, Energies 13:6512. https://doi.org/10.3390/en13246512
22. Garip Z, Ekinci E, Alan A (2023) Day-ahead solar photovoltaic energy forecasting based on weather data using LSTM networks: a comparative study for photovoltaic (PV) panels in Turkey. Electr Eng. https://doi.org/10.1007/s00202-023-01883-7
23. Chen C, Duan S, Cai T, Liu B (2011) Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol Energy 85:2856–2870. https://doi.org/10.1016/j.solener.2011.08.027
24. Azizi N, Yaghoubirad M, Farajollahi M, Ahmadi A (2023) Deep learning based long-term global solar irradiance and temperature forecasting using time series with multi-step multivariate output. Renew Energy 206:135–147. https://doi.org/10.1016/j.renene.2023.01.102
25. An W, Zheng L, Yu J, Wu H (2022) Ultra-short-term prediction method of PV power output based on the CNN–LSTM hybrid learning model driven by EWT. J Renew Sustain Energy 14:053501. https://doi.org/10.1063/5.0101922
26. Jiang Y, Zheng L, Ding X (2021) Ultra-short-term prediction of photovoltaic output based on an LSTM-ARMA combined model driven by EEMD. J Renew Sustain Energy 13:046103. https://doi.org/10.1063/5.0056980
27. Karout Y, Thil S, Eynard J, Guillot E, Grieu S (2023) Hybrid intrahour DNI forecast model based on DNI measurements and sky-imaging data. Sol Energy 249:541–558. https://doi.org/10.1016/j.solener.2022.11.032
28. Visser L, AlSkaif T, Hu J, Louwen A, van Sark W (2023) On the value of expert knowledge in estimation and forecasting of solar photovoltaic power generation. Sol Energy 251:86–105. https://doi.org/10.1016/J.SOLENER.2023.01.019
29. Kumari P, Toshniwal D (2021) Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. Appl Energy 295:117061. https://doi.org/10.1016/J.APENERGY.2021.117061
30. Lee D, Kim K (2019) Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. Energies 12:215. https://doi.org/10.3390/EN12020215
31. Li P, Zhou K, Lu X, Yang S (2020) A hybrid deep learning model for short-term PV power forecasting. Appl Energy 259:114216. https://doi.org/10.1016/J.APENERGY.2019.114216
32. Ghimire S, Deo RC, Raj N, Mi J (2019) Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Appl Energy 253:113541. https://doi.org/10.1016/J.APENERGY.2019.113541
33. Wang F, Yu Y, Zhang Z, Li J, Zhen Z, Li K (2018) Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting. Appl Sci 8:1286. https://doi.org/10.3390/APP8081286

34. Suresh V, Janik P, Rezmer J, Leonowicz Z (2020) Forecasting solar PV output using convolutional neural networks with a sliding window algorithm. Energies 13:723. https://doi.org/10.3390/EN13030723
35. Wang K, Qi X, Liu H (2019) A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Appl Energy 251:113315. https://doi.org/10.1016/J.APENERGY.2019.113315
36. Wojtkiewicz J, Hosseini M, Gottumukkala R, Chambers TL (2019) Hour-ahead solar irradiance forecasting using multivariate gated recurrent units. Energies 12:4055. https://doi.org/10.3390/EN12214055
37. Husein M, Chung IY (2019) Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: a deep learning approach. Energies 12:1856. https://doi.org/10.3390/EN12101856
38. Agga A, Abbou A, Labbadi M, El Houm Y, Ou Ali IH (2022) CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electr. Power Syst. Res. 208:107908. https://doi.org/10.1016/J.EPSR.2022.107908
39. Zang H, Liu L, Sun L, Cheng L, Wei Z, Sun G (2020) Short-term global horizontal irradiance forecasting based on a hybrid CNN-LSTM model with spatiotemporal correlations. Renew Energy 160:26–41. https://doi.org/10.1016/j.renene.2020.05.150
40. Elsaraiti M, Merabet A (2022) Solar power forecasting using deep learning techniques. IEEE Access 10:31692–31698. https://doi.org/10.1109/ACCESS.2022.3160484
41. Qu J, Qian Z, Pei Y (2021) Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. Energy 232:120996. https://doi.org/10.1016/J.ENERGY.2021.120996
42. Alaraj M, Kumar A, Alsaidan I, Rizwan M, Jamil M (2021) Energy production forecasting from solar photovoltaic plants based on meteorological parameters for qassim region, Saudi Arabia. IEEE Access 9:83241–83251. https://doi.org/10.1109/ACCESS.2021.3087345
43. Gandoman FH, Abdel Aleem SHE, Omar N, Ahmadi A, Alenezi FQ (2018) Short-term solar power forecasting considering cloud coverage and ambient temperature variation effects. Renew Energy 123:793–805. https://doi.org/10.1016/J.RENENE.2018.02.102
44. Ahamed MS, Guo H, Tanino K (2021) Cloud cover-based models for estimation of global solar radiation: a review and case study. Int J Green Energy 19:175–189. https://doi.org/10.1080/15435075.2021.1941043
45. Sarkar MNI (2016) Estimation of solar radiation from cloud cover data of Bangladesh. Renewables Wind Water, Sol 3:11. https://doi.org/10.1186/s40807-016-0031-7
46. Zhang W, Chen X, He K, Chen L, Xu L, Wang X, Yang S (2022) Semi-asynchronous personalized federated learning for short-term photovoltaic power forecasting. Digit Commun Networks. https://doi.org/10.1016/J.DCAN.2022.03.022
47. Mellit A, Pavan AM, Lughi V (2021) Deep learning neural networks for short-term photovoltaic power forecasting. Renew Energy 172:276–288. https://doi.org/10.1016/J.RENENE.2021.02.166
48. Mustaqeem, Ishaq M, Kwon S (2022) A CNN-Assisted deep echo state network using multiple Time-Scale dynamic learning reservoirs for generating Short-Term solar energy forecasting. Sustain Energy Technol Assessments 52:102275. https://doi.org/10.1016/J.SETA.2022.102275
49. Wang K, Qi X, Liu H (2019) Photovoltaic power forecasting based LSTM-Convolutional Network. Energy 189:116225. https://doi.org/10.1016/j.energy.2019.116225
50. Sharma N, Mangla M, Yadav S, Goyal N, Singh A, Verma S, Saber T (2021) A sequential ensemble model for photovoltaic power forecasting. Comput Electr Eng 96:107484. https://doi.org/10.1016/j.compeleceng.2021.107484

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Heat and Mass Transfer
https://doi.org/10.1007/s00231-024-03495-8

RESEARCH

A novel investigation on single-loop pulsating heat pipe filling with hybrid nanofluids: numerical and experimental

Prem shanker yadav[1,2] · Jitendra sharma[1] · Mohd Hussain[1] · Imran ahmed Khan[1] · Kartik Goyal[1] · Samer Fikry Ahmed[3]

Received: 31 January 2024 / Accepted: 1 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
Pulsating heat pipes (PHPs) employ to multiphase heat transfer between condensers and evaporators. The efficacy of PHP is predominantly contingent upon the thermos-physical property exhibited by its working fluid. The exergy analysis of evaporator, adiabatic and condenser system were performed to evaluate the efficiency and sustainability of energy conversion processes, it was also seen that exergy loss of condenser was 1.01% higher than evaporator. Further, 2-D numerical simulation of a cryogenic pulsating heat pipe (CPHP) was conducted and also numerical and experimental simulation was conducted, where simulation results agreed with experimental results with 10% similarity. The simulation employs the volume of fluid (VOF) model to capture the dynamics of two-phase liquid–vapor flow within the CPHP employing liquid acetone – Al_2O_3 as the working fluid. The diameter of the single turn is systematically varied, ranging from 1 mm to 2.5 mm, while maintaining the filling ratio (FR) within the range of 25% to 75%. The evaporator temperature is adjusted within the span of 85 K to 115 K. The PHP exhibited best thermal performance at inner diameter of 2 mm among (1, 1.5, 2, 2.5 mm) and filling ratio of 55% among (25%, 45%, 55%, 65% and 75%).

Abbreviations

PHP	Pulsating Heat Pipe
CLPHP	Closed loop pulsating heat pipe
Al_2O_3	Aluminum oxide
BP	Boiling point
LHv	Latent heat of vaporization
D	Diameter
ρ_l	Density of liquid
ρ_v	Density of vapor
µm	Micro - meter
d_{crit}	Critical diameter
T_{max}	Maximum Temperature
Cp	Specific heat capacity
$T_{e,o}$	Temperature of working fluid exiting evaporator
σ	Boltzman constant
CFD	Computational Fluid Dynamics
d	Internal diameter
σ	Surface Tension
FR	Filling ratio
g/s	gram/second
mm	millimeter
MPa	Mega Pascal
nm	Nano - meter
VOF	Volume of fluid
Te	Evaporator Temperature
Tsat	Saturated Temperature
Te,i	Temperature of working fluid before entering evaporator
Tc,i	Temperature of working fluid before entering condenser
ε	Emissivity
Evap	Evaporator

1 Introduction

The Pulsating Heat Pipe (PHP), a passive heat transmission system, has been utilized for approximately 30 years. In regards to its structure, the PHP can be described as a hollow and winding pipe. Khandekar and Groll [1] is credited with the development of these devices, which exhibit tremendous

✉ Prem shanker yadav
prem0jss@gmail.com

[1] Department of Mechanical Engineering, JSS Academy of Technical Education, Noida, India

[2] Department of Mechanical Engineering, Delhi Technological University, Delhi, India

[3] Mechanical and Industrial Engineering Department, College of Engineering, Qatar University, Doha, Qatar

Published online: 10 July 2024

potential in various applications such as solar energy utilization, waste heat recovery, aircraft thermal management, and electronics cooling as shown in Fig. 1 [2]. PHPs are passive cooling systems that operate with two-phase flow, providing rapid heat transfer. They consist of 3—sections: the adiabatic, evaporator section and the condenser. The condenser as well as evaporator parts are thermally connected to a cold sink and a heat source respectively, using thermal anchors. Capillary tubing is employed to connect the condenser and evaporator part, these are bent back and forth repeatedly. There are three primary types of PHPs: Closed-loop with a check valve, Closed Loop, and Closed-ended PHPs [3]. In a CLPHP, a slug flow pattern is observed, characterized by the presence of vapor and liquid slugs, as depicted in Fig. 1. As vapor plugs move towards the condenser section, they undergo a reduction in volume caused by re-condensation and temperature influences, with certain vapor bubbles fully transitioning into the liquid phase. Boiling in the evaporator region results in the growth of existing vapor bubbles and the formation of new ones. These expansions and contractions of the two-phase fluid generate pressure fluctuations necessary for driving the flow. Fluids characterized by a high $(dP/dT)_{sat}$ are considered preferable due to the substantial pressure shift observed in vapor bubbles with only a slight temperature adjustment at saturation levels [4].

However, the selection of the working fluid plays a crucial role in determining the performance of a PHP. Selecting a suitable working fluid that meets the operating temperature range, heat transfer requirements, and compatibility with the materials used in the PHP can be a challenge. Furthermore, the availability and cost of the chosen fluid can also impact

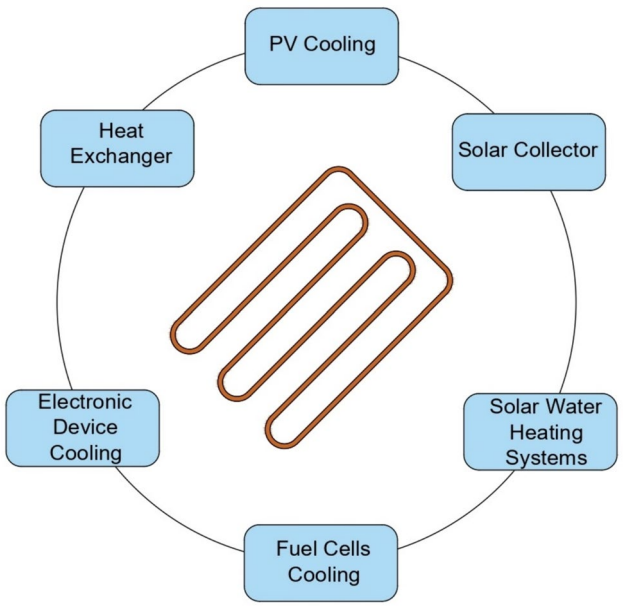


Fig. 1 Applications of PHP

the practicality of PHP implementation [5, 6]. Further, Zhang et al. [7] examined that it is simpler to start up a working fluid with low surface tension, low dynamic viscosity and high thermal conductivity. Acetone's low latent heat and low surface tension make it simpler to dry out than other working fluids [8]. The increased flow speed of nanofluids in PHP was noted by Zhang et al. [9]. PHP can typically start and run between 20 and 80% of the filling ratio. While, high filling ratio prevents the production of vapor plugs and impedes flow, whereas a low filling ratio makes it difficult to cause oscillation of the operational fluid [10, 11].

The sensible heat also plays a crucial role in dissipating a significant amount of heat as the heat flux supplied to the evaporation section is minimal. In such cases, where the heat transfer rate is relatively small, the majority of heat transfer occurs through the sensible heat mechanism. Further, the ability to maintain the oscillation motion at a lower required heat flux is crucial for the performance and effectiveness of the PHP. Therefore, opting for a working fluid with lower viscosity can contribute to enhancing the overall heat transfer efficiency of the system [12, 13]. The results of Qu et al. [14] demonstrated that among the tested working fluids such as acetone, methanol and water, acetone exhibited the best heat transfer performance in the PHP system. Despite the low heat flux input, acetone proved to be the most effective choice in terms of facilitating heat transfer within the PHP. The major portion of heat transfer is accomplished through the sensible heat of the liquid, while the latent heat of vaporization plays a secondary role [15]. Liang et al. [16] investigated the thermal performance of a cryogenic PHP and specifically examined the influence of condenser temperature and filling ratio. The findings demonstrated that a lower filling ratio led to a dry-out phenomenon occurring at lower heat inputs. Zhang et al. [17] conducted experiments taking FC-72, ethanol, and water as working fluid, and found that the best fill ratio for all three fluids was around 70%, with lower and higher fill ratios resulting in lower thermal performance. At this fill ratio, the mean temperatures in the heating section were lowest and the thermal performance was highest. In an experimental study conducted by Saha et al. [18], two open PHPs were examined. The results of the experiment indicated that the PHP with a 0.9 mm inner diameter exhibited superior performance compared to the PHP with a 1.5 mm inner diameter. It is worth noting that the physical properties of the working fluid can influence the impact of the inner diameter on the PHP's performance, as highlighted by Rittidech et al. [3]. Specifically, it was observed that a larger inner diameter resulted in improved heat transfer performance for R123, while it led to poorer performance for ethanol. Hybrid nanofluids, which are a combination of base fluids and nanoparticles, are being

increasingly explored for use in PHPs due to their potential to enhance thermal performance. The addition of nanoparticles to the base fluid in PHPs can offer several benefits such as improved thermal conductivity and heat transfer coefficient, reduced thermal resistance and controlled fluid flow. However, there are some challenges in utilization of nanofluids such as particle agglomeration, difficulty in nanoparticle dispersion [19–21].

Karthikeyan et al. [22] explored the use of nano-fluids, specifically Ag-deionized water and Cu-deionized water, within the range of heating power from 50 to 240 W. Their findings indicated a noteworthy 33.3% increase in the heat transfer capacity of the PHP when employing nano-fluids compared to the utilization of pure deionized water. Ji et al. [23] scrutinized the heat transfer characteristics of the PHP utilizing Al_2O_3-water with varying particle sizes (80 nm, 50 nm, 22 µm, and 2.2 µm). Their observations revealed that the initiation power of the PHP decreased proportionally with the reduction in particle size. Tanshen et al. [24] examined the thermal resistance of the PHP employing nano-fluids containing multi-walled carbon nanotubes (MWCNT) across four mass ratios (0.2, 0.3, 0.05, and 0.1 wt%). The outcomes they obtained suggested that the PHP attained its peak inner evaporation pressure, and concurrently, the thermal resistance of the PHP reached its minimum at a mass ratio of 0.2 wt%. Changes in boiling performance in heat pipes directly affect the surface properties, as indicated in relevant literature sources [25–27]. The presence of nanoparticles in the base fluid does not significantly increase the thermal conductivity of the working fluid. However, the oscillatory movement of particles within the working fluids may offer additional enhancement to the thermal performance of the CLPHP.

From this comprehensive review, it becomes evident that there are two significant research gaps noticed.

(a) No previous studies have explored the effects of acetone- Al2O3 on the exergy and sustainability. These analyses serve as contemporary tools for evaluating the effectiveness of systems, offering valuable insights and conclusions.
(b) Additionally, there is a notable absence of research investigating the optimum diameter and filling ratio for acetone- Al2O3 fluid to get maximum performance.

The first section of the study evaluates the thermodynamic analysis and sustainability analysis. The subsequent section outlines the governing equations and computational techniques utilized for the CFD analysis at various diameter to analyze the flow characteristics within a PHP. Getting optimum diameter, filling ratio was varied to get the enhanced performance of PHP.

2 Experimental setup and description of operation of a PHP

The internal pipe diameter, represented by the Bond dimensionless number ($B_0 = (\rho_l - \rho_v)gd/\sigma$) plays a crucial role in characterizing phase separation within the PHP device [28]. In Taylor bubble-rising experiments involving water boiling in vertical tubes, it has been observed that when the bond number is less than 2, the vapor bubble and the wall film maintain radial symmetry due to the dominant influence of surface tension over gravitational force [29]. PHPs exhibit remarkable operational stability across various orientations, allowing them to function successfully. The stability is attributed to the diverse flow regimes that can be observed in a PHP, as depicted in Fig. 2. Under low heat loads, the vapor phase diffuses into the liquid phase, leaving the nucleation sites on the walls. As evaporation intensifies, the number of small and micro-bubbles increases, with the smaller bubbles migrating towards the core region. These bubbles engage in interactions with each other and frequently merge, forming

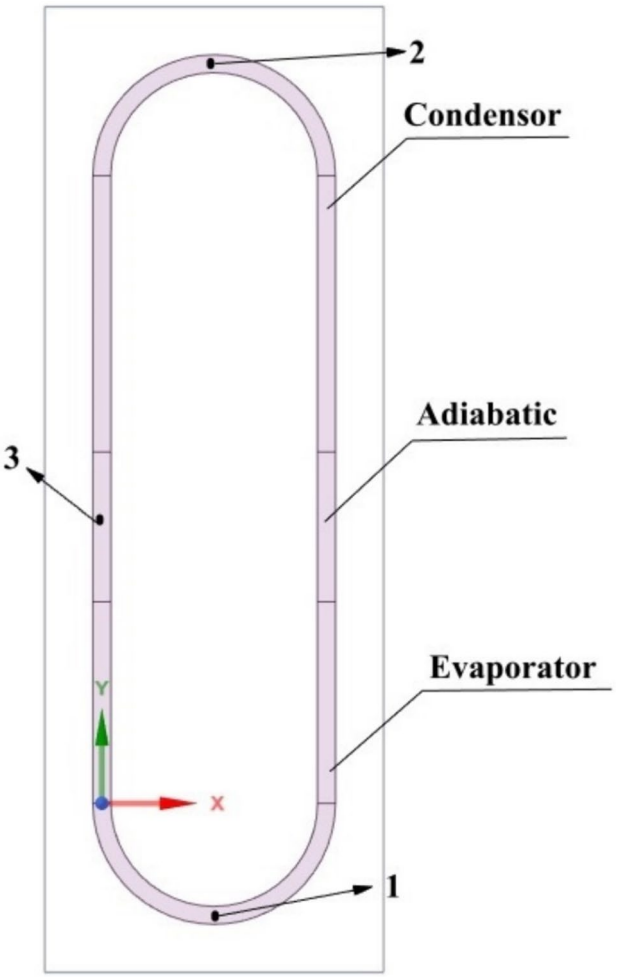


Fig. 2 Single loop pulsating heat pipe model

larger bubbles. With a continued rise in heat load, the dimensions of these aggregations of bubbles progressively enlarge, eventually resulting in the separation of the mixed-phase fluid into alternating liquid segments known as slugs and vapor segments referred to as plugs. When the capillary length is larger than internal diameter of the pipe, the aforementioned formations spread radially and occupy the pipe's core space. With increased energy input, the phase change intensifies the turbulence and movement, making the segments unstable. In such unstable conditions, the structures have a tendency to break down into chaotic swirls, eventually forming small droplets. This phenomenon gives rise to a flow regime with enhanced heat transmission capabilities, known as annular flow [30, 31]. Within such a regime, the flow field within The PHP can be segmented into two clearly defined regions. The core region comprises a liquid film wall layer along with a mixture of dispersed liquid and vapor phase. Notably, significant interactions occur between the wall and core regions in this mode. At extremely elevated levels of core flux, droplets carried along in the core flow have the potential to condense into liquid masses, akin to vapor conglomerates. This condensation process contributes to pressure instability [32]. PHPs leverage these instabilities to generate a unique mixed-phase flow, setting them apart from conventional heat pipes. The fluid within the PHP is propelled along the serpentine tube by a pressure gradient, which arises due to local variations in the rates of condensation and evaporation. Pulsation is initiated by the shifting direction of chains of slugs and plugs due to an imbalance in the total force [33]. The areas of a PHP have varying velocities due to expansion from liquid evaporation and contraction from vapor condensation, respectively. These localized motion fields have the potential to interact with each other as waves and change in strength as a result.

A region exhibiting alike flow velocity and direction can form if certain locations are connected as the heat load rises. The circular flow regime, which enhances heat capacity and can withstand a large heat load, tends to emerge in the single-directional field [34]. The robust interactions of the wall boundary layer also contribute to heat transmission. A bubble traveling through the channel exhibits an unbalanced head and a less orderly tail. The bubble having thin liquid film is disrupted by the turbulence created in the wake zone, leading to a local increase in the heat transfer coefficient. This enhanced convection is primarily responsible for thermal transfer that surpasses conduction within the solid wall by orders of magnitude [35].

Figure 3 illustrates the experimental apparatus employed in the present study. The condensation, adiabatic and evaporation sections comprise a length of 50 mm each. The pulsating heat pipe is constructed using quartz glass having 6 mm outer diameter. For visualization purposes, it has an inner diameter of 3.8 mm. The current PHP has a sufficiently small inner diameter, allowing the working fluid, water, to spontaneously disperse as liquid–vapor slugs. It is oriented vertically, with the bottom-heating pattern. A heating wire, measuring 0.4 mm in diameter, is wrapped around the surface of the evaporation section. Additionally, a water cooler is submerged in the condensation area. The water cooler's inlet temperature is maintained at 25 °C, while a fix flow rate of 4.69 g/s is prescribed. The current PHPs undergo evacuation to 0.079 MPa using a vacuum pump. Subsequently, water is introduced into the PHPs through the injection entrance, as illustrated in Fig. 3. In this study, different filling ratios of 30%, 40%, 50%, and 60% are employed for the PHPs. Furthermore, the heating power can be adjusted within the range of 5W to 40W by modifying the output voltage. The experimental setup depicted in Fig. 3 includes six T-type thermocouples positioned on the outer wall of the PHP. These thermocouples are specifically labeled as Te_1 and Te_2 for the evaporation section, Ta_1 and Ta_2 for the adiabatic section, and Tc_1 and Tc_2 for the condensation section. Real-time temperature data is collected using an agilent 34970A data acquisition system, which is connected to a computer and programmed to scan the temperatures at intervals of 5 s. This continuous data acquisition enables the determination of the experimental thermal resistance (Rex) of the PHP. The average temperatures of the evaporation and condensation sections are denoted as T_e and T_c, respectively. The input heating power is represented by Q in the equation $Rex = \frac{(T_e - T_c)}{Q}$.

The working fluids having minimum latent heat of vaporization, lower boiling point and liquid specific heat can begin pulsating very readily inside the pulsating heat pipe, however, they are low energy carrier. Further, low dynamic viscosity fluids experience lesser flow resistance resulting strong pulsating effects due to high flow rate [33]. Table 1 shows the thermos-physical properties of water, acetone and acetone-Al_2O_3 and it was observed that acetone-Al2O3 consist of low dynamic viscosity, liquid specific heat and latent heat of vaporization and that improves pulsating effects in PHPs. The diameter is varied from 1 mm to 2.5 mm based on bond no which shows the critical diameter 2.1 mm. Further, investigation is performed order to get the optimum filling ratio which is considered from 25 to 75%.

Uncertainty analysis involves estimating the true value from the mean of repeated measurements. The errors observed in the test data primarily stem from the heating procedure and the subsequent recording of temperature data. For a given parameter, six experiments were conducted, and the average was used for error analysis. The uncertainty associated with the total thermal resistance of PHP can be estimated using Holman's square root rule as shown in Eq. 1 [36].

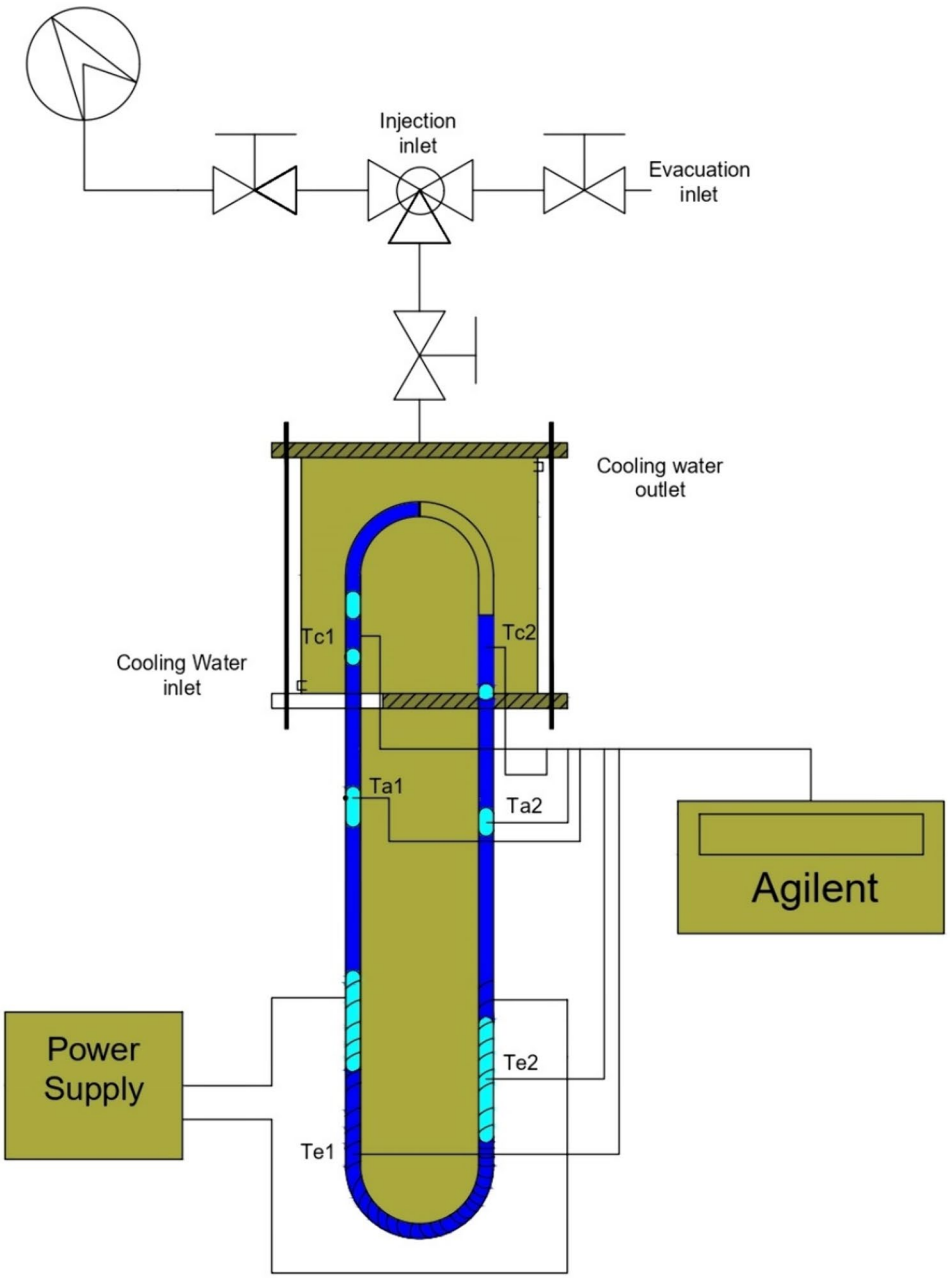


Fig. 3 Experimental set-up single loop PHP

$$U_E = \{[(\frac{\partial E}{\partial X_1})u_1]^2 + (\frac{\partial E}{\partial X_2})u_2]^2 + (\frac{\partial E}{\partial X_3})u_3]^2 + \ldots\ldots\ldots + (\frac{\partial E}{\partial X_n})u_n]^2\}^{1/2} \quad (1)$$

Here, E denote the fundamental variable impacting the other variables, $X_1,\ldots,Xn$. X corresponds to the quantities utilized during the experimental process, and $u_1,\ldots,u_n$

Table 1 Effect of Al_2O_3 on thermo-physical properties of working fluid

Working fluids	LHv (Hrg)	Surface Tension ($\sigma \times 10^3$) N/m	Boiling point (Ts ^{0}C)	Liquid density, ρ_l (kg/m^3 20^{0}c)	Thermal conductivity, λ_l (W/(Mk)) 20^{0}c	Liquid specific heat, C_{pl} (KJ/kgK) 20^{0}C	Dynamic viscosity ($\upsilon_l \times 10^3$) Pa.s 20^{0}C	$(dp/dt)^*_{sat}$ X 10^3 (Pa/K, 80^{0}C)	Ref.
Water	2257	72.8	100	998	0.599	4.18	1.01	1.92	[33]
Acetone	523	23.7	56.2	792	0.170	2.35	0.32	6.27	[33]
Acetone-Al_2O_3	516	24.8	-	798	0.206	2.32	0.28	-	Present

represent the uncertainties associated with these measurements. Consequently, the comprehensive uncertainty of the experimental findings can be described as following Eq. (2).

$$\frac{\Delta R}{R} = \sqrt{\left(\frac{\Delta(T_{ave,e} - T_{ave,c})}{Q_{in}}\right)^2} = \pm 1.96\% \tag{2}$$

The change in thermal resistance (ΔR), when considered, reveals that the maximum uncertainty in the overall heat input and thermal resistance in PHP falls well within the ± 1.96% range.

The total uncertainty value is ± 1.96%, which falls within the range reported by prior studies [37].

3 Exergy analysis

This is a thermodynamic approach employed to evaluate the quality or usefulness of energy within a system. It takes into account both the energy content and the thermodynamic irreversibility in the system [38]. In the case of a PHP, two-phase heat transfer passive device, it can be performing an exergy analysis to evaluate the efficiency and losses within the system. Here a single turn PHP with an adiabatic, evaporator, and a condenser section. The exergy balance equations for each section are as follows:

Evaporator section

The heat transfer in a PHP involves two main components: liquid slug flow and vapor flow. The exergy change in the evaporator section is given by Eq. 3 [3]:

$$\Delta Ex_E = mC_p[(Te, i - Te, o) - To \ln\left(\frac{Te, i}{Te, o}\right)\}] \tag{3}$$

The heat loss can be expressed as the sum of convective heat transfer and radiative heat transfer as in Eqs. 4 and 5 [39]

$$\Delta Ex_{e,1} = \Delta Ex_{co,1} + \Delta Ex_{ra,1} \tag{4}$$

$$\Delta Ex_{co,1} = hco \times As \times \Delta T \tag{5}$$

where, $\Delta Ex_{e,}$ $\Delta Ex_{e,l}$, $\Delta Ex_{co,l}$,, $\Delta Ex_{ra,l}$ are net exergy change across evaporator, exergy loss through evaporator, convective exergy loss and radiative exergy loss respectively.

Radiative heat transfer in a PHP evaporator occurs through thermal radiation between the evaporator surface and the surroundings. The radiative heat transfer rate can be estimated using the following Eq. 6 [39]

$$\Delta Ex_{ra,1} = \sigma \times \varepsilon \times As \times (T_{sf}^4 - T_{sr}^4) \tag{6}$$

Adiabatic section

Since the adiabatic section does not involve any heat transfer, the exergy change is only due to the fluid friction losses:

$$\Delta Ex_adiabatic = W_friction$$

where:

$\Delta Ex_adiabatic$ is the change in exergy in the adiabatic section,

$W_friction$ is the work done against fluid friction losses. It is assumed that the pressure drop is very less throughout the adiabatic section therefore the losses in adiabatic section is negligible.

Condenser section

Similar to the evaporator section, the exergy change in the condenser section can be expressed as per Eq. 7 [7]:

$$\Delta Ex_c = mC_p[(Tc, i - Tc, o) - To \ln\left(\frac{Tc, i}{Tc, o}\right)\}] \tag{7}$$

where:

The heat loss can be expressed as the sum of convective heat transfer and radiative heat transfer as in Eqs. 8 and 9 [9].

$$\Delta Ex_{c,1} = \Delta Ex_{co,1} + \Delta Ex_{ra,1} \tag{8}$$

$$\Delta Ex_{co,1} = hco \times Asc \times \Delta T \tag{9}$$

where, ΔEx_c $\Delta Ex_{c,1}$, $\Delta Ex_{co,1}$, $\Delta Ex_{ra,1}$ are net exergy change across condenser, exergy loss through condenser, convective exergy loss and radiative exergy loss respectively. Radiative heat transfer in a PHP evaporator occurs through thermal radiation between the evaporator surface and the surroundings. The radiative heat transfer rate can be estimated using the following Eq. 10 [40]

$$\Delta Ex_{ra,1} = \sigma \times \varepsilon \times Asc \times (T_{sf}^4 - T_{sr}^4) \tag{10}$$

The exergy efficiency of a PHP can be calculated by comparing the actual exergy transfer in the system to the maximum possible exergy transfer. The exergy efficiency $\eta_{\Delta E_X}$ is defined as the ratio of the actual exergy transfer to the maximum exergy transfer as given in Eq. 11 [39, 40]:

$$\eta_{\Delta E_x} = \frac{(Ex_out - Ex_in)}{Ex_in_\max} \tag{11}$$

Sustainability Assessment Sustainable development requires the resources to be used efficiently. Employing the exergy analysis method proves to be a valuable tool for optimizing advantages and utilizing resources efficiently. In this analysis, the SI method directly related to exergy efficiency $\eta_{\Delta E_x}$ is used to improve and contribute to the sustainable development as follows [41] as in Eq. 12:

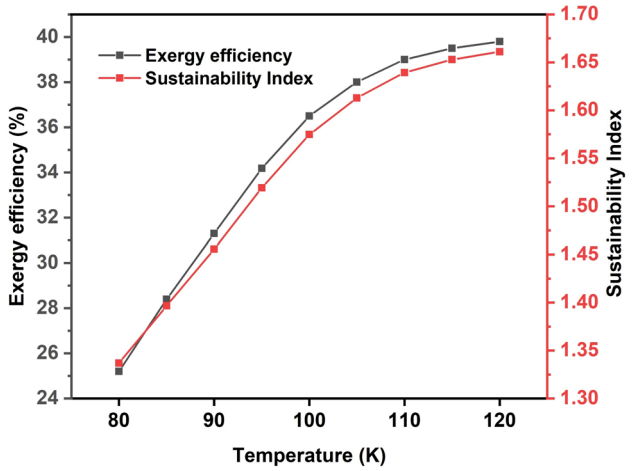


Fig. 4 Exergy efficiency and sustainability vs temperature

$$SA = \frac{1}{(1 - \eta_{\Delta EX})} \tag{12}$$

Figure 4 demonstrate the sustainability index and exergy efficiency in terms of reference temperature. It can be observed that exergy efficiency initially increased and further rate of increment got decrease. This is due to rise in exergy loss. In the given range of reference temperature, the exergy efficiency is highest at 120 K of temperature.

The sustainability index also showed the similar pattern of exergy efficiency and that confirms the resources can be utilized efficiently.

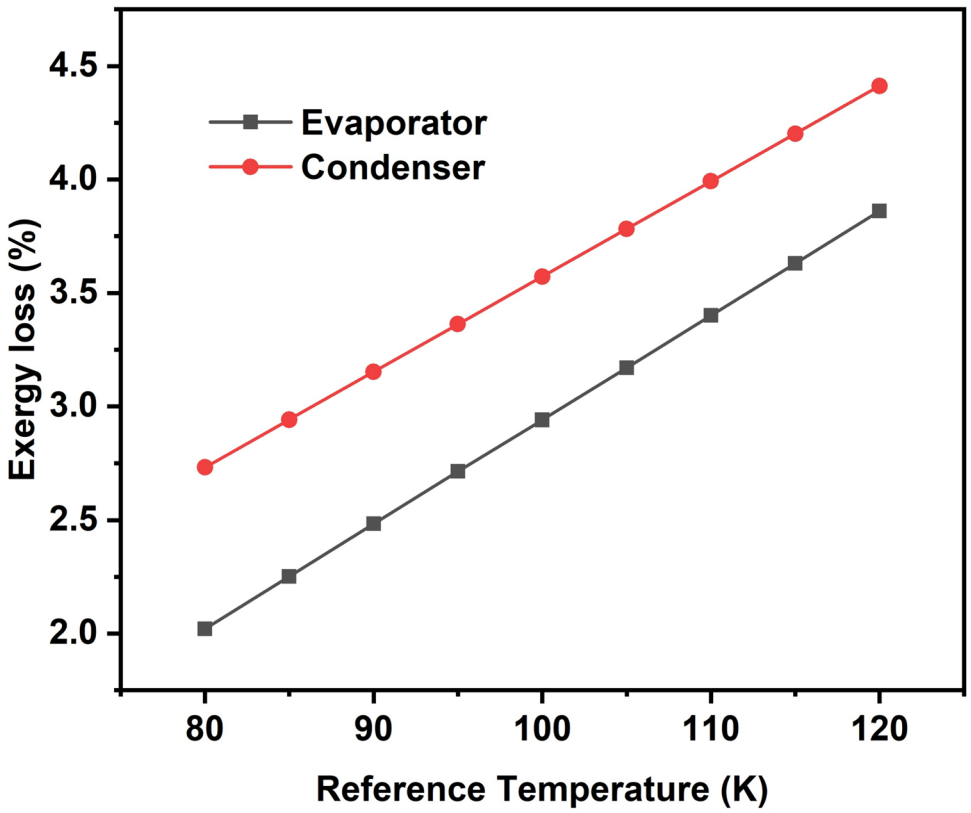
In the evaporator, as the reference temperature rises, the temperature difference between the working fluid and the surroundings also increases. This larger temperature gradient results in higher heat transfer irreversibilities and, consequently, higher exergy losses as shown in Fig. 5.

Similarly, in the condenser, a higher reference temperature means a higher temperature difference between the working fluid and the surroundings during heat rejection. This increased temperature gap results in greater irreversibilities and higher exergy losses.

4 Simulation and modeling

Figure 6 illustrates a 2-D physical model of a single loop PHP, where acetone – Al_2O_3 serves as the operating fluid. The PHP is positioned for bottom heating, and it has an overall length span of 57 mm. In order to ensure that the dominance of surface tension force over gravitational force is evident within the channel, which measures 1 mm, the calculation of the inner critical diameter of the PHP involves the use of the Bond number (Eq. 1, The inner diameter (D) of the CPHP must be smaller than the critical diameter (D_c) for proper functionality). Thus, the current CPHP model's inner diameter is set at 1 mm.

Fig. 5 Exergy loss vs reference temperature



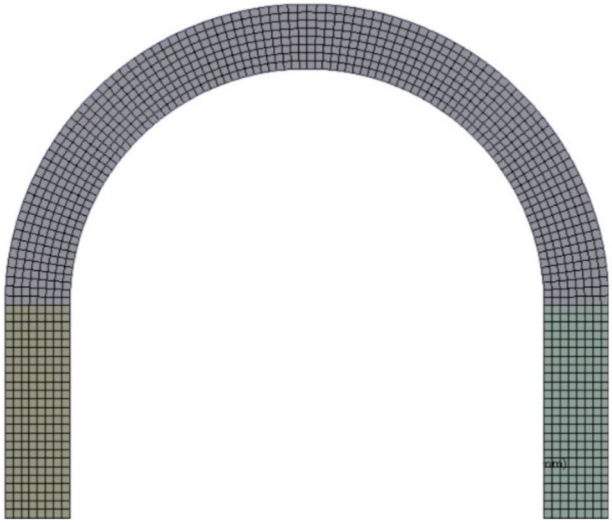


Fig. 6 A mesh structure adopted in a 2-D CPHP model

To accurately account for the surface tension effect, a quadrilateral mesh is favored over tetrahedral and triangular meshes [42]. Accordingly, quadrilateral structured mesh has been employed. Figure 6 illustrates the utilized quadrilateral mesh in the 2-D PHP model. The thickness of the first layer adjacent to the wall is set at 0.18 mm.

The flow dynamics in a PHP can be described as a two-phase gas-liquid flow regime. This regime is characterized by the presence of gas bubbles interspersed with liquid slugs [43]. The VOF approach enables the modeling of the two-phase flow by tracking the interface between the gas and liquid phases. The VOF approach ensures that the distinct gas and liquid phases within a PHP, characterized by bubbles and liquid slugs, are appropriately represented without mixing or overlapping. The variables are integrated for each introduced phase in the model, ensuring that the total sum of volume fractions for all phases within a computational cell equals unity. The properties and variables in a cell exhibit either two phase mixture or single phase and that relies on volume fraction values (Table 2).

Consider the fluid volume fraction in a cell, denoted as αq, representing the volume percentage of the fluid labeled as "qth.". There are three possible conditions: $\alpha q = 1$, $\alpha q = 0$ and $0 < \alpha q < 1$ are indicated as full cell, empty cell (for qth fluid) and interface between two fluids. Depending on the value of αq, appropriate variables and properties are assigned to every control volume within the domain. Further, it is postulated that phase transition transpires at the saturation temperature when transitioning between the vapor and liquid phases. Equation (13) represents mass conservation equation [44].

$$\frac{\partial(\alpha_v \rho_v)}{\partial t} + \nabla.(\alpha_v \rho_v \upsilon_v) = m_{lv} - \dot{m}_{vl} \tag{13}$$

The interphase mass transfer is described as in Eqs. 14 and 15 [44].

Here, $T_l > T_{sat}$ (for evaporation)

$$\dot{m}_{lv} = r_{lv}.\alpha_l \rho_l \frac{(T_l - T_{sat})}{T_{sat}} \tag{14}$$

If, $T_{sat} > T_v$ (for condensation)

$$\dot{m}_{vl} = r_{vl}.\alpha_l \rho_l \frac{(T_l - T_{sat})}{T_{sat}} \tag{15}$$

where, r indicates relaxation factor,

The relaxation factor, denoted as 'r,' is set to a default value of 0.1 s-1 to ensure the interface temperature remains in proximity to the saturation temperature of the working fluid [45]. Throughout the domain, the momentum equation (Eq. 16) is computed, the momentum equation's outcome is influenced by the volume fractions of each phase [46].

$$\frac{\partial(\rho\vec{v})}{\partial t} + \nabla.(\rho\vec{v}\vec{v}) = -\nabla P + \nabla.[\mu(\nabla\vec{v} + \nabla\vec{v}^T)] + \rho\vec{g} + \vec{F}_{vol} \tag{16}$$

Surface tension emerges due to the cohesive forces among molecules present in the fluid. These surface forces become dominant, particularly in mini channels. The CSF model is employed for surface tension modelling in FLUENT. Incorporating surface tension into the VOF formulation introduces it as a source term in the momentum equation. The volume force acts on the surface are given by Eq. (17) [47].

$$F_{vol} = \sigma_{lv} \frac{\alpha_v \rho_v k_l \nabla\alpha_l + \alpha_l \rho_l k_v \nabla\alpha_v}{\frac{1}{2}(\rho_v + \rho_l)} \tag{17}$$

Table 2 Numerical simulation matrix

Items	Content	Items	Content
Multiphase model	VOF Implicit /Explicit	Volume Fraction	Geo-Reconstruct
Viscous model	k-epslion	Model	2D transient
Energy and Momentum and	Upwind second Order	Mesh	Quadrilateral
Pressure interpolation	PRESTO	First layer thickness	0.18 mm
Gravitational acceleration	-9.81 m/s 2, Y axis	Modeling of surface tension force	CSF

The curvature (k) is mathematically represented as follows: $k_l = \frac{\nabla \alpha_l}{\nabla \alpha_v}$, $k_v = \frac{\nabla \alpha_v}{\nabla \alpha_l}$.

Further, the Eq. (18) shows the energy equation contributed among phases [48]. Where, Sh exhibits the energy source due to phase change. It is acquired by multiplying the latent heat by the rate of mass transfer such as

$$S_h = -h_{LH}\, m_{lv} = h_{LH}\, \dot{m}_{vl}$$

$$\frac{\partial}{\partial t}(\rho E) + \nabla.(\vec{v}(\rho E + P)) = \nabla.(K.\nabla T + (\overline{\overline{\tau}}.\vec{v})) + S_h \tag{18}$$

The two terms of Eq. (18) in right side are the sum of energy transfer through viscous dissipation and conduction.

In Eq. (19), the temperature (T) and energy (E) are considered as mass-averaged variables in volume of fluid (VOF) model [47].

$$E = \frac{\alpha_v \rho_v E_v + \alpha_l \rho_l E_l}{\alpha_v \rho_v + \alpha_l \rho_l} \tag{19}$$

where, Ev and E_l are depends on specific heat of phases and shared temperature as in Eqs. 20 and 21.

$$E_v = C_{v,v}(T - T_{sat}) \tag{20}$$

$$E_l = C_{v,l}(T - T_{sat}) \tag{21}$$

The properties K, ρ and μ are common to all phases, as presented below:

$$K = \alpha_v K_V + \alpha_l K_l,\ \rho = \alpha_v \rho_v + \alpha_l \rho_l,\ \mu = \alpha_v \mu_v + \alpha_l \mu_l$$

5 Results and discussion

5.1 Validation of model and grid sensitivity test

The results of laboratory research on a PHP were compared with previous simulations to verify the accuracy of the current numerical method [28]. The geometry used in the laboratory investigation is completely consistent with that of the numerical study, with a filling ratio of 55% and having ethylene glycol as working fluid. The thermal resistance measured in the laboratory study differs from that in the current numerical study by less than 5% as shown in Fig. 7. The concordance between the simulation and experiment findings shows that the current numerical simulations can accurately represent the two-phase liquid/vapor flow in the oscillatory heat pipe. Where Q and Te represents the heat input and evaporator temperature.

Moreover, based on our current understanding, the quantity of grids exerts a somewhat indirect influence

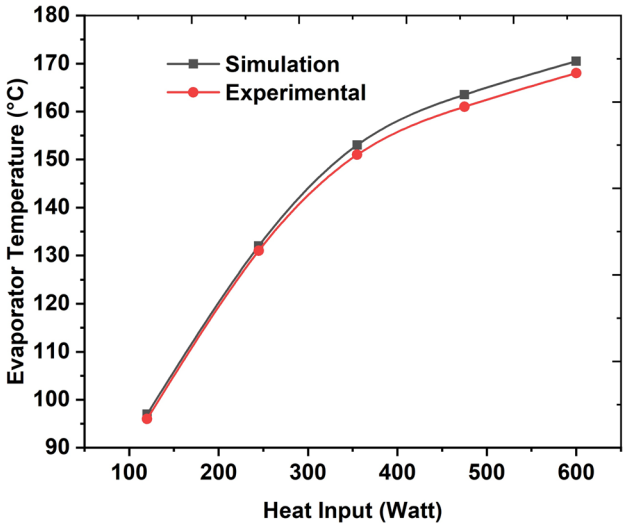


Fig. 7 Fluid temperature vs time (Evaporator and condenser) at FR (42%)

on the outcomes of numerical simulations. Four distinct grid numbers— 75,623, 55,545, 22,323 and 38,432— are employed to assess grid independence and meet the requisite criteria. Throughout the testing, all parameters remain constant, except for the grid count. When utilizing grid numbers 75623 and 55,545, as depicted in Fig. 8, the simulation results demonstrate nearly identical consistency, with a relative variance of approximately 1%. Consequently, due to the numerical precision and computational efficiency, grid number 55545 is selected for use in this study.

5.2 Effect of internal diameter

The viscous force and surface tension are directly affected by internal diameter of PHP. Where, bond no limits the

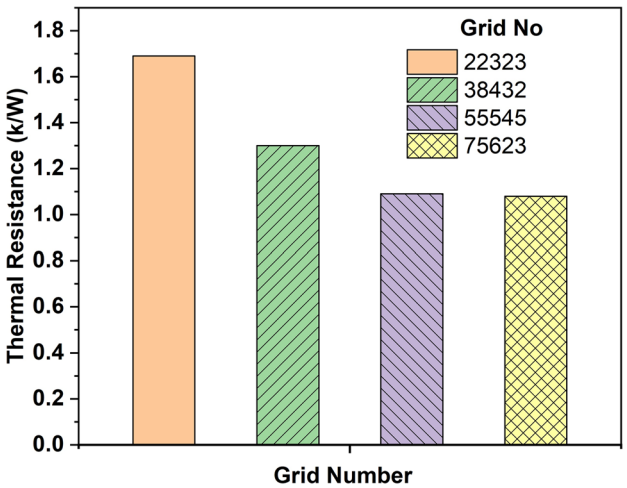


Fig. 8 Grid independence test (Q = 45 W)

maximum diameter for sustained liquid slugs [1]. The critical diameter of a pulsating heat pipe is presented in Eq. 22

$$d_{crit} = 2\sqrt{\frac{\sigma}{g(\rho_l - \rho_v)}} \tag{22}$$

The critical diameter for a pulsating heat pipe (PHP) with acetone-Al_2O_3 as the working fluid can be calculated using the eq. 2. The critical diameter comes out to be 2.1 mm. When the diameter of the PHP (Pulsating Heat Pipe) channels is smaller, the capillary forces tend to be stronger. As a result, there is an increase in liquid-vapor phase separation and the formation and growth of bubbles are promoted. This phenomenon has been distinctly observed in the PHP illustrated in the Figs. (9 and 10). When the channel diameter is closer to the critical diameter (D_{crit}), the PHP exhibits better slug-plug characteristics, leading to an improvement in overall performance. However, the minimum diameter to which a PHP can be reduced depends on several factors, including the properties of the working fluid, surface tension, fluid viscosity, channel geometry, and the desired operating conditions but there is no universal minimum diameter for a PHP [49] (Fig. 9).

Though, the figures show that the benefit of reducing the total thermal resistance, achieved by expanding the range of movement, is maintained up to a 1 mm internal diameter. The longer periods of standstill indicate a decline in surface tension forces beyond this point. The most noticeable difference is that the tiny internal diameter has more short bubbles than the large internal diameter. However, the smaller diameter leads to increased thermal resistance within the PHP, as heat transfer area decreases while the flow resistance increases as shown in Fig. 10. The rationale of above is also suggested by study [50] (Fig. 11).

It was also observed that capillary became more influential with decrease in internal diameter. A decrease in diameter enhanced capillary action due to increased surface tension effects, which may aid in overcoming gravitational forces, thus facilitating better fluid circulation. In contrast to other inner tube sizes, the lower thermal resistance value was also found for an inner diameter of 2 mm and this result is align with studies [51].

5.3 Effect of filling ratio

The transient fluid temperature variation as well as pressure drop with time in evaporator for FR 25%-75% at evaporator temperature Te = 105 K is shown in Figs. (11, 12, 13, 14, 15, 16, 17, 18, 19 and 20). It's also presumed that this filling ratio remains consistent over time. and flow is incompressible since small temperature variation. Figure 12 illustrates the T_{max} graph acquired at 0.25 FR. The oscillations observed in T_{max} over time exhibit a similar nature across various filling ratios. Yet, deviating from the norm, an intriguing trend emerges for a filling ratio of 0.55, as depicted in Fig. 13. It was observed that At a filling ratio of 25%, there exists a greater abundance of vapor plugs within the evaporator. Therefore, the time

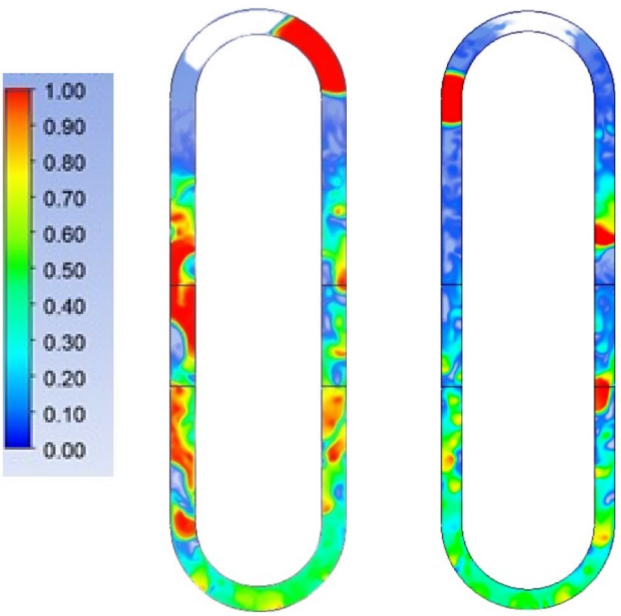


Fig. 9 Volume liquid fraction vs time value (D: 2.5 mm and 2 mm) (left to right)

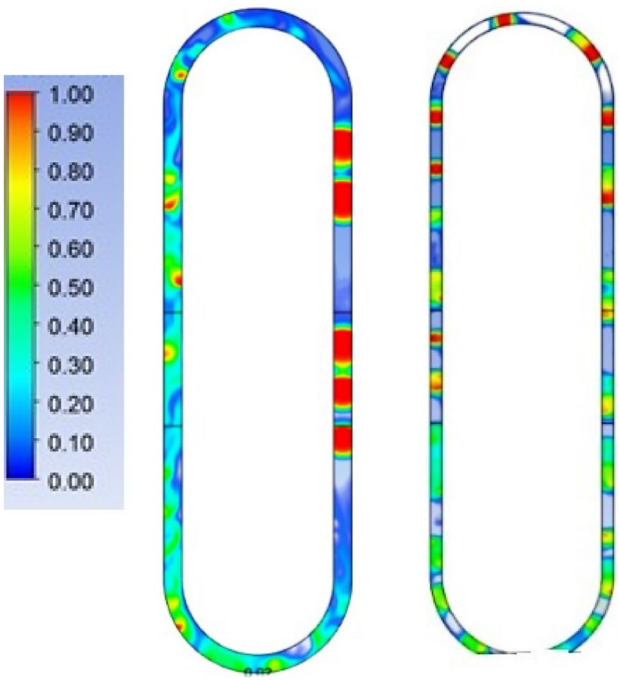


Fig. 10 Volume liquid fraction vs time value (D: 1.5 mm and 1 mm) (left to right)

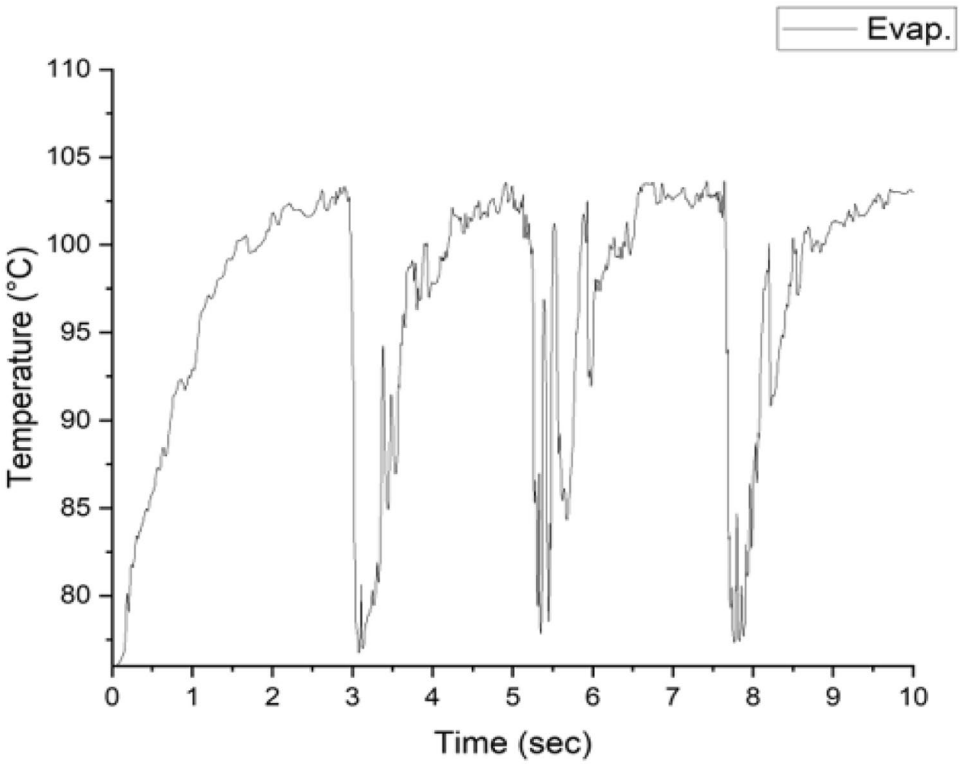


Fig. 11 Temperature Vs Time graph for Fill ratio 25%

averaged temperature of the acetone-Al_2O_3 in the evaporator section is notably elevated. In Figs. 12, 13 and 14, with the increase in filling ratio from 45 to 75%, there is a corresponding decrease in the time-averaged temperature difference of the working fluid. This phenomenon is primarily attributed to the higher.

quantity of liquid slugs presents at increased filling ratios (45% to 75%), accompanied by a higher circulation velocity of the slug flow. Since, the filling ratio determines the amount of working fluid available for heat transfer (Fig. 15).

A higher filling ratio allows for a larger amount of working fluid, resulting in higher heat transfer capability. This is because a larger volume of working fluid can absorb more heat before reaching a critical temperature [2]. It is worth to note that the temperature gradient with time decreased with increasing filling ration.

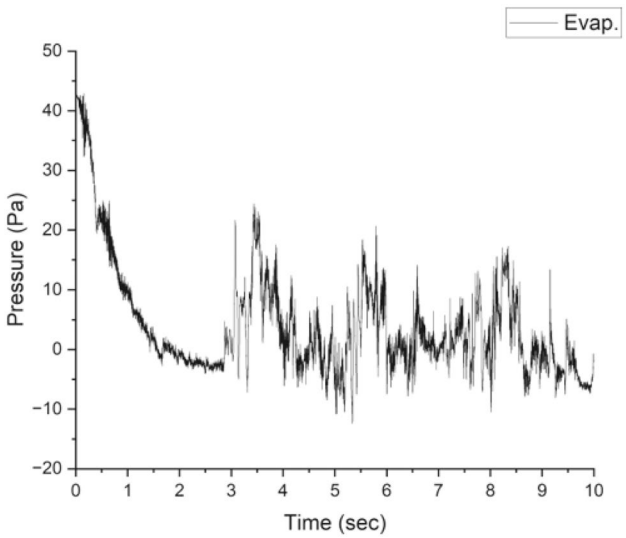


Fig. 12 Pressure vs time graph for fill ratio 25%

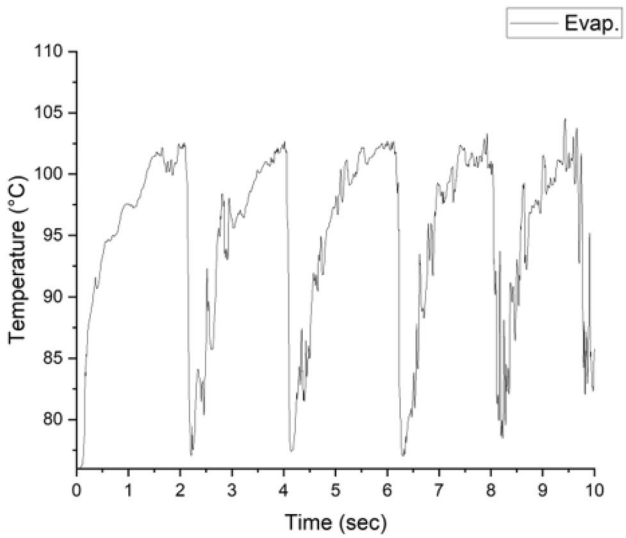


Fig. 13 Temperature vs time graph for Fill ratio 45%

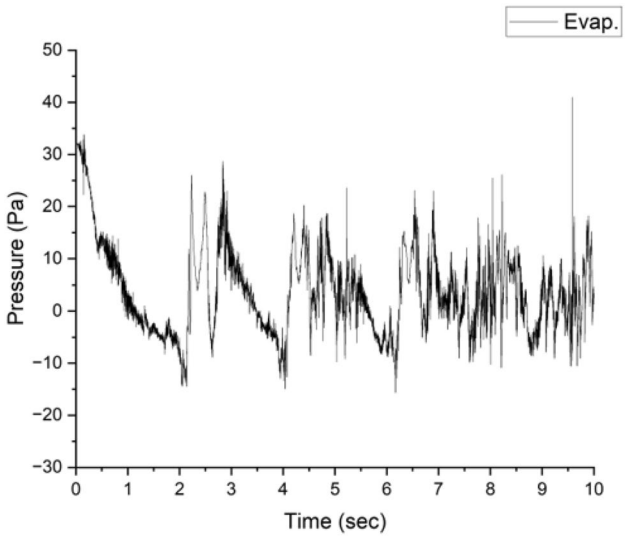


Fig. 14 Pressure vs time graph for fill ratio 45%

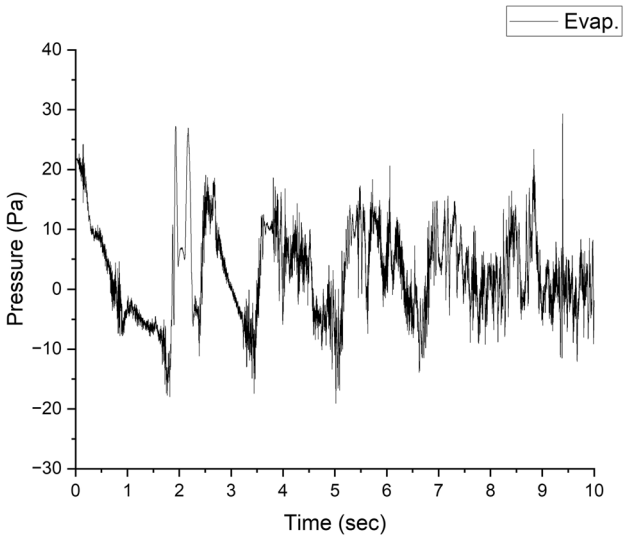


Fig. 16 Pressure vs time graph for fill ratio 55%

However, after FR = 55%, the maximum temperature also increased. At a constant filling ratio, the temperature gradient of the working fluid (the temperature difference between condenser and evaporator) rises proportionally with the improve in heat input. Nevertheless, when the filling ratio was elevated at a constant heat load, the fluid temperature gradient reduced. Hence at high heat input situation, high FR is more suitable to avoid dry out in a PHP (Fig. 16).

As the liquid content within the PHP steadily diminishes (accompanied by a corresponding increase in vapor volume), the heat transfer experiences an initial boost owing to the heightened movement of vapor plugs. Nonetheless, concurrently, the quantity of sensible heat transfer undergoes a continuous decline as a result of the diminishing liquid content. Therefore, an optimum point exists at a specific filling ratio (here, FR = 55%). The rationale of above study is also suggested by [52] Since, PHPs rely on capillary forces to transport the working fluid between the condenser and evaporator section (Fig. 17).

Capillary forces are generated by the interaction between the working fluid and the inner walls of the tube. A lower filling ratio may result in reduced capillary forces due to a smaller amount of fluid available for generating these forces [53] (Fig. 18).

Figures 12, 13 and 14 shows the pressure drop with time, and was observed that filling ratio having 75% exhibited on average highest pressure drop among all filling ratio. This is due to higher filling ratios generally lead to higher pressure

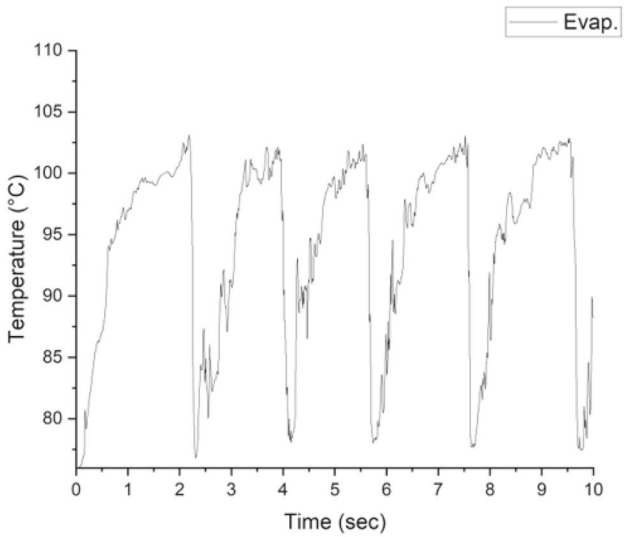


Fig. 15 Temperature vs time graph for fill ratio 55%

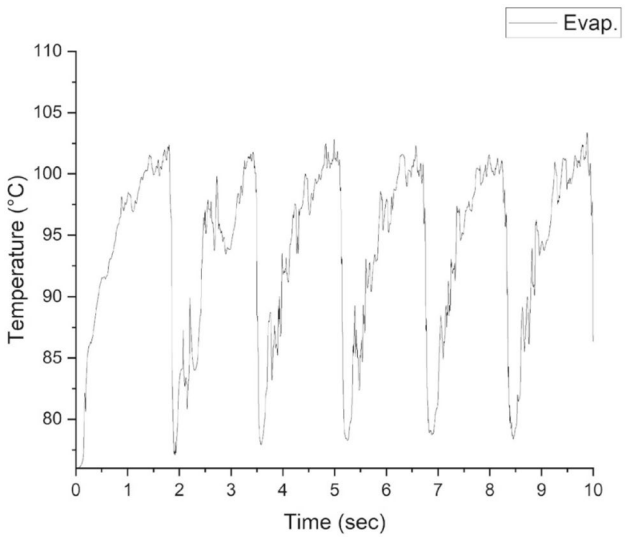


Fig. 17 Temperature vs time graph for fill ratio 65%

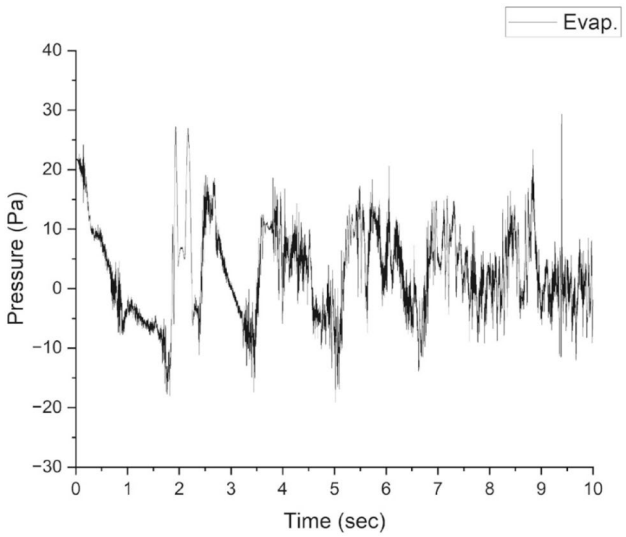


Fig. 18 Pressure vs time graph for fill ratio 65%

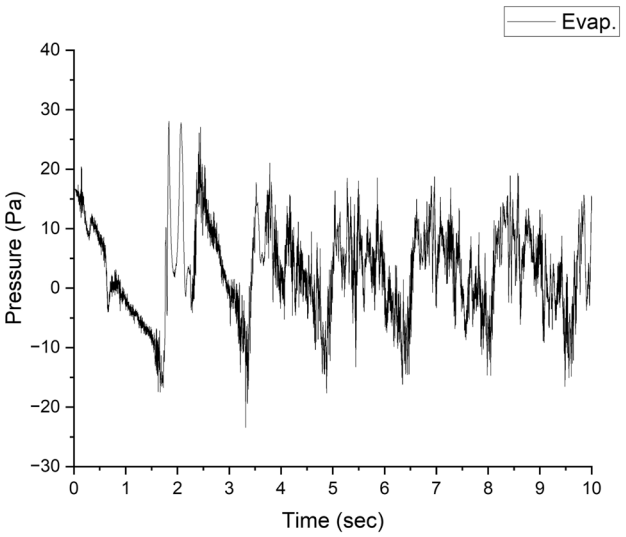


Fig. 20 Pressure vs time graph for fill ratio 75%

drops due to increased frictional resistance between the working fluid and the tube walls. This can impact the fluid flow rate and the overall system performance [54]. However, maximum pressure drop was seen for 25% filling ratio at initially small duration. It occurs due to gas entrainment: The gas plugs formed in slug flow can get entrained with the liquid slugs (Fig. 19).

As the PHP operates, these gas plugs can become dislodged from the liquid slugs, creating gas bubbles that travel through the system. The movement of gas bubbles along with the liquid slugs contributes to the overall pressure drop [55] and frictional effects: Slug flow introduces a high interfacial area between the liquid and gas phases. This increased interfacial area leads to increased frictional effects as the liquid and gas flow past each other. Frictional pressure losses occur due to the shearing forces between the two phases, resulting in additional pressure drop [56]. It is worth to note that the time averaged working fluid temperature difference (thick straight lines) evaporator decreases as FR is raised from 25 to 75%.

Additionally, there is an increased number of vapor plugs within the evaporator depicted in Fig. 20 when FR = 25%. Consequently, the time-averaged temperature of the working fluid within the evaporator section is significantly elevated as there in [57]. The time-averaged temperature differential of the working fluid decreases in Figs. 15, 16, 17, 18, 19, and 20 as FR increases from 45 to 75%. This shift is primarily attributable to the greater presence of liquid slugs at higher filling ratios (45% to 70%) and increased circulation velocities of slug flow.

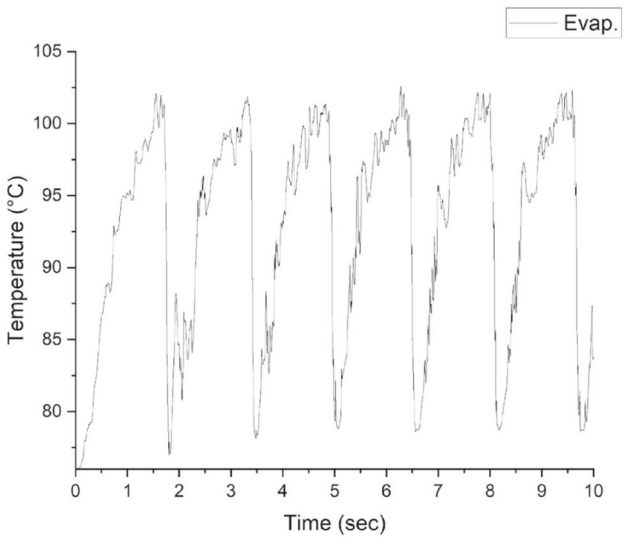


Fig. 19 Temperature vs time graph for fill ratio 75%

6 Conclusion

The pulsating heat pipe utilizing an Acetone – Al_2O_3 mixture, single turn tube, and cylindrical geometry demonstrated successful operation across various diameters and fill ratios under closed PHP valve conditions. The following conclusions are drawn from the study.

1. Increasing the diameter in a pulsating heat pipe involves a trade-off. While it can enhance heat transfer area and fluid flow dynamics, there's a limit beyond which capillary forces may become less effective. Therefore, PHP showed highest performance at 2 mm internal diameter among (1 mm, 1.5 mm, 2 mm and 2.5 mm)

2. In mixtures, the inhibition of phase transitions can significantly delay the onset of dry-out in the PHP, even when operating at low filling ratios (25%, 45%), while, Due to the reduced flow rate, the heat transmission capabilities of the PHP with acetone-Al2O3 mixtures are less effective at higher filling ratios (65 and 75%).
3. As the temperature increased, the exergy efficiency also increased linearly. However, it became nearly constant at higher temperature. Sustainability index also showed similar trends to exergy efficiency.
4. Exergy loss was lower for evaporator than condenser for a range of temperature. This is due to low temperature difference in evaporator, which can contribute to lower exergy losses.

7 Future scope

Investigate the implementation of optimal PHP configurations in real-world scenarios, considering external factors such as fluctuations in temperature. Evaluate the reliability and enduring stability of continuously operating PHP systems. Extend the investigation to explore the potential for enhancing optimization through the utilization of diverse nanoparticles and alternative working fluids. Enhance the practical applicability of PHP technology by assessing its scalability and adaptability across various industrial sectors.

Acknowledgements We would like to thank Prof. Zafar Said, University of Sharjah for key insite

Author contributions J,M,I,K,P,S: wrote the main manuscript text, prepared figures and tables P,S: reviewed the manuscript

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

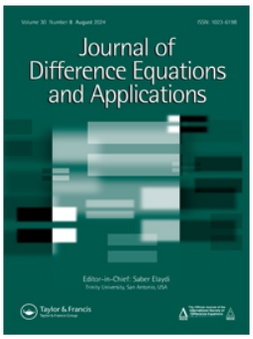
References

1. Khandekar S, Groll M (2003) On the definition of pulsating heat pipes: An overview. In Proc. 5th Minsk Int. Conf. (Heat Pipes, Heat Pumps and Refrigerators), Minsk, Belarus 707–719
2. Han X, Wang X, Zheng H, Xu X, Chen G (2016) Review of the development of pulsating heat pipe for heat dissipation. Renew Sustain Energy Rev 59:692–709
3. Rittidech S, Terdtoon P, Murakami M, Kamonpet P, Jompakdee W (2003) Correlation to predict heat transfer characteristics of a closed-end oscillating heat pipe at normal operating condition. Appl Therm Eng 23(4):497–510
4. Aminian E, Kamali M, Vatanjoo E, Saffari H (2022) Theoretical analysis on condensation heat transfer on microstructured hybrid hydrophobic-hydrophilic tube. Heat Mass Transf 58(7):1207–1221
5. Sedighi M, Ayoobi A, Aminian E (2024) Numerical investigation on forced convection enhancement within an oil cooler through the simultaneous use of porous media and nanofluid. Proceed Inst Mech Eng Part E J Proc Mech Eng 238(1):331–341
6. Zhang D, Wang L, Xu B, Li Q, Wang S, An Z (2024) Experimental and simulation study on flow heat transfer characteristics of flat pulsating heat pipe with wide and narrow interphase channels. Appl Therm Eng 122806
7. Zhang D, He Z, Jiang E, Shen C, Zhou J (2021) A review on start-up characteristics of the pulsating heat pipe. Heat Mass Transf 57:723–735
8. Pachghare PR, Mahalle AM (2014) Thermo-hydrodynamics of closed loop pulsating heat pipe: An experimental study. J Mech Sci Technol 28:3387–3394
9. Zhang D, He Z, Guan J, Tang S, Shen C (2022) Heat transfer and flow visualization of pulsating heat pipe with silica nanofluid: An experimental study. Int J Heat Mass Transf 183:122100
10. Faghri A (2012) Review and advances in heat pipe science and technology. J Heat Transfer 134(12):123001
11. Markal B, Aksoy K (2021) The combined effects of filling ratio and inclination angle on thermal performance of a closed loop pulsating heat pipe. Heat Mass Transf 57:751–763
12. Talesh Bahrami HR, Aminian E, Saffari H (2020) Energy transfer enhancement inside an annulus using gradient porous ribs and nanofluids. J Energy Res Technol 142(12):122102
13. Arab M, Soltanieh M, Shafii MB (2012) Experimental investigation of extra-long pulsating heat pipe application in solar water heaters. Exp Thermal Fluid Sci 42:6–15
14. Qu W, Yang B 2009 Performances of flat plate pulsating heat pipes. In International Conference on Micro/Nanoscale Heat Transfer 43918:403–410
15. Aminian E, Saffari H (2022) Experimental analysis of dropwise condensation heat transfer on a finned tube: Impact of pitch size. Proceed Inst Mech Eng Part A J Power Energy 236(4):752–759
16. Liang Q, Li Y, Wang Q (2018) Effects of filling ratio and condenser temperature on the thermal performance of a neon cryogenic oscillating heat pipe. Cryogenics 89:102–106
17. Zhang XM (2004) Experimental study of a pulsating heat pipe using FC-72, ethanol, and water as working fluids. Experimental Heat Transfer 17(1):47–67
18. Saha M, Feroz CM, Ahmed F, Mujib T (2012) Thermal performance of an open loop closed end pulsating heat pipe. Heat Mass Transf 48:259–265
19. Siavashi M, Bahrami HT, Aminian E, Saffari H (2019) Numerical analysis on forced convection enhancement in an annulus using porous ribs and nanoparticle addition to base fluid. J Central South Univ 26(5):1089–1098
20. Smrity AMA, Yin P (2024) Design and performance evaluation of pulsating heat pipe using metallic nanoparticles based hybrid nanofluids. Int J Heat Mass Transf 218:124773
21. Babaei A, Aminian E, Saffari H (2023) Numerical study of heat transfer and pressure drop of nanofluids in a combined porous media of hydrophobic and hydrophilic surfaces. Proceed Inst Mech Eng Part A J Power Energy 237(6):1308–1328
22. Karthikeyan VK, Ramachandran K, Pillai BC, Solomon AB (2014) Effect of nanofluids on thermal performance of closed loop pulsating heat pipe. Exp Thermal Fluid Sci 54:171–178
23. Ji Y, Ma H, Su F, Wang G (2011) Particle size effect on heat transfer performance in an oscillating heat pipe. Exp Thermal Fluid Sci 35(4):724–727
24. Tanshen MR, Munkhbayar B, Nine MJ, Chung H, Jeong H (2013) Effect of functionalized MWCNTs/water nanofluids on thermal resistance and pressure fluctuation characteristics in oscillating heat pipe. Int Commun Heat Mass Transfer 48:93–98

25. Bhuwakietkumjohn N, Rittidech S (2010) Internal flow patterns on heat transfer characteristics of a closed-loop oscillating heat-pipe with check valves using ethanol and a silver nano-ethanol mixture. Exp Thermal Fluid Sci 34(8):1000–1007
26. Qu J, Wu H (2011) Thermal performance comparison of oscillating heat pipes with SiO2/water and Al_2O_3/water nanofluids. Int J Therm Sci 50(10):1954–1962
27. Jouhara H, Khordehgah N, Almahmoud S, Delpech B, Chauhan A, Tassou SA (2018) Waste heat recovery technologies and applications. Therm Sci Eng Prog 6:268–289
28. Rudresha S, Babu ER, Thejaraju R (2023) Experimental investigation and influence of filling ratio on heat transfer performance of a pulsating heat pipe. Therm Sci Eng Prog 38:101649
29. White ET, Beardmore RH (1962) The velocity of rise of single cylindrical air bubbles through liquids contained in vertical tubes. Chem Eng Sci 17(5):351–361
30. Khandekar S (2004) Thermo-hydrodynamics of closed loop pulsating heat pipes
31. Mameli M, Manno V, Filippeschi S, Marengo M (2014) Thermal instability of a closed loop pulsating heat pipe: Combined effect of orientation and filling ratio. Exp Thermal Fluid Sci 59:222–229
32. Vatanjoo E, Aminian E, Kamali M, Saffari H (2023) Theoretical analysis on condensation heat transfer on the hydrophobic–hydrophilic hybrid surfaces with the impact of the Marangoni convection. Proceed Inst Mech Eng Part E J Process Mech Eng 237(3):658–672
33. Wang W, Cui X, Zhu Y (2017) Heat transfer performance of a pulsating heat pipe charged with acetone-based mixtures. Heat Mass Transf 53:1983–1994
34. Mameli M, Mangini D, Vanoli GFT, Araneo L, Filippeschi S, Marengo M (2016) Advanced multi-evaporator loop thermosyphon. Energy 112:562–573
35. Mameli M, Marengo M, Khandekar S (2014) Local heat transfer measurement and thermo-fluid characterization of a pulsating heat pipe. Int J Therm Sci 75:140–152
36. Holman JP (2001) Experimental methods for engineers
37. Yadav PS, Said Z, Gautam R, Raman R, Caliskan H (2023) Novel investigation on atomization, performance, and emission characteristics of preheated jatropha oil methyl ester and ethyl ester. Energy 270:126870
38. Rashidi MM, Mahariq I, Alhuyi Nazari M, Accouche O, Bhatti MM (2022) Comprehensive review on exergy analysis of shell and tube heat exchangers. J Therm Anal Calorim 147(22):12301–12311
39. Dincer I, Rosen MA (2012) Exergy: Energy, environment and sustainable development. Newnes
40. Bejan A, Tsatsaronis G, Moran MJ (1995) Thermal design and optimization. John Wiley & Sons
41. Rosen MA, Dincer I, Kanoglu M (2008) Role of exergy in increasing efficiency and sustainability and reducing environmental impact. Energy Policy 36(1):128–137
42. Pastukhov VG, Maydanik YF (2016) Development of a pulsating heat pipe with a directional circulation of a working fluid. Appl Therm Eng 109:155–161
43. Karthikeyan VK, Ramachandran K, Pillai BC, Brusly Solomon A (2015) Understanding thermo-fluidic characteristics of a glass tube closed loop pulsating heat pipe: Flow patterns and fluid oscillations. Heat Mass Transf 51:1669–1680
44. Wang J, Xie J, Liu X (2020) Investigation of wettability on performance of pulsating heat pipe. Int J Heat Mass Transf 150:119354
45. Lin Z, Wang S, Shirakashi R, Zhang LW (2013) Simulation of a miniature oscillating heat pipe in bottom heating mode using CFD with unsteady modeling. Int J Heat Mass Transf 57(2):642–656
46. Minko KB, Artemov VI, Klementiev AA (2024) CFD simulation of the inundation effect for saturated propane vapor condensation on the surface of a horizontal tube using the volume of fluid method. Int Commun Heat Mass Transfer 154:107432
47. Aminian E, Moghadasi H, Saffari H (2020) Magnetic field effects on forced convection flow of a hybrid nanofluid in a cylinder filled with porous media: A numerical study. J Therm Anal Calorim 141:2019–2031
48. Liu Y, Dan D, Wei M, Zheng S, Sun J (2024) Numerical investigation on the start-up and heat transfer performance of dual-diameter pulsating heat pipes. Appl Therm Eng 236:121709
49. Li Z, Sarafraz MM, Mazinani A, Moria H, Tlili I, Alkanhal TA, . . . Safaei MR (2020) Operation analysis, response and performance evaluation of a pulsating heat pipe for low temperature heat recovery. Energy Convers Manag 222:113230
50. Wang J, Pan Y, Liu X (2021) Investigation on start-up and thermal performance of the single-loop pulsating heat pipe with variable diameter. Int J Heat Mass Transf 180:121811
51. Babu ER, Reddy NC, Babbar A, Chandrashekar A, Kumar R, Bains PS, . . . Ammarullah MI (2024) Characteristics of pulsating heat pipe with variation of tube diameter, filling ratio, and SiO2 nanoparticles: Biomedical and engineering implications. Case Stud Therm Eng 55:104065
52. Jajarm ARA, Goshayeshi HR, Bashirnezhad K, Chaer I, Toghraie D, Salahshour S (2024) Combined effect of the magnetic field, orientation, and filling ratio on cylindrical pulsating heat pipe using distilled water and distilled water/Fe3O4 nanofluid. J Magn Magn Mater 590:171712
53. Spinato G, Borhani N, Thome JR (2015) Understanding the self-sustained oscillating two-phase flow motion in a closed loop pulsating heat pipe. Energy 90:889–899
54. Chen X, Lin Y, Shao S, Wu W (2019) Study on heat transfer characteristics of ethane pulsating heat pipe in middle-low temperature region. Appl Therm Eng 152:697–705
55. Rittidech S, Donmaung A, Kumsombut K (2009) Experimental study of the performance of a circular tube solar collector with closed-loop oscillating heat-pipe with check valve (CLOHP/CV). Renew Energy 34(10):2234–2238
56. Zolfaghari A, Aminian E, Saffari H (2022) Numerical investigation on entropy generation in the dropwise condensation inside an inclined pipe. Heat Transfer 51(1):551–577
57. Barua H, Ali M, Nuruzzaman M, Islam MQ, Feroz CM (2013) Effect of filling ratio on heat transfer characteristics and performance of a closed loop pulsating heat pipe. Proc Eng 56:88–95

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



Journal of Difference Equations and Applications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gdea20

A uniformly convergent method for two-parameter singularly perturbed parabolic partial differential equations with a large shift

Aditya Kaushik, Nitika Sharma, Manju Sharma, Monika Arora & Shivani Jain

To cite this article: Aditya Kaushik, Nitika Sharma, Manju Sharma, Monika Arora & Shivani Jain (09 Jul 2024): A uniformly convergent method for two-parameter singularly perturbed parabolic partial differential equations with a large shift, Journal of Difference Equations and Applications, DOI: 10.1080/10236198.2024.2376565

To link to this article: https://doi.org/10.1080/10236198.2024.2376565

Published online: 09 Jul 2024.

Submit your article to this journal

Article views: 9

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=gdea20

JOURNAL OF DIFFERENCE EQUATIONS AND APPLICATIONS
https://doi.org/10.1080/10236198.2024.2376565

Taylor & Francis
Taylor & Francis Group

Check for updates

A uniformly convergent method for two-parameter singularly perturbed parabolic partial differential equations with a large shift

Aditya Kaushik[a], Nitika Sharma[a], Manju Sharma[b,c], Monika Arora[d] and Shivani Jain[a]

[a]Department of Applied Mathematics, Delhi Technological University, Delhi, India; [b]Department of Mathematics, Indraprastha College for Women, University of Delhi, Delhi, India; [c]Department of Mathematics, KVA DAV College, Karnal, India; [d]Department of Mathematics, Miranda House, University of Delhi, Delhi, India

ABSTRACT

A class of singularly perturbed two-parameter parabolic partial differential equations with a large shift is studied. These equations include time-dependent variables and exhibit sensitivity to small perturbations in the system parameters. Moreover, the shift captures the influence of neighbouring states or events on the current evolution of the system. It adds memory-like behaviour to the problem, making their analysis and numerical solution more intricate. The solution of these equations exhibits a multiscale character. The interior and boundary layers exist across which the solution has a steep gradient. A higher-order accurate numerical method is proposed on a non-uniform mesh to solve the problem. The method combines an upwind difference scheme in space and a Crank-Nicolson scheme in time. The method is unconditionally stable, and parameters uniformly convergent with second-order accuracy in time and first-order accuracy in the space variable. Besides, the paper presents numerical results and illustrations to support theoretical estimates.

ARTICLE HISTORY
Received 6 March 2024
Accepted 30 June 2024

KEYWORDS
Singular perturbation; two-parameter problem; non-uniform mesh; boundary/interior layers

2020 MATHEMATICS SUBJECT CLASSIFICATIONS
65M06; 65Q10; 65Q20

1. Introduction

The problem we consider involves a singularly perturbed parabolic equation with two small parameters which control the system's behaviour and a large shift. This type of problem forms an essential basis for several physical [17], biological and chemical processes [18], including chemical flow [8], lubrication theory [10], and reactor theory [24], to name a few among numerous others. The presence of two small parameters leads to interesting phenomena and challenges in the analysis and numerical solution of these problems. The solution exhibits twin boundary layers due to the presence of two parameters and an interior layer due to the large shift. O'Malley has started the study of singular perturbation problems [23–25]. He has shown that the nature of the problem and the occurrence of boundary layers in the solution is affected by the magnitude of the two parameters involved. The nature of these layers depends on the ratio of μ^2 and ε [26].

CONTACT Manju Sharma manjusharma@ip.du.ac.in Department of Mathematics, Indraprastha College for Women, University of Delhi, Delhi 110054, India

© 2024 Informa UK Limited, trading as Taylor & Francis Group

Many researchers have worked to develop uniformly convergent numerical methods for the solution of two-parameter singularly perturbed parabolic problems [4,5,7,32]. In [21], the researchers consider a two-parameter singularly perturbed time-delay parabolic equation. They employ a fitted operator finite difference scheme to approximate the numerical solution. The first step involves discretising the time variables using the Crank-Nicolson method. The semidiscrete problem is further discretized in space using the exponentially fitted tension-spline finite difference method. The study reveals that the numerical scheme exhibits almost second-order uniform convergence in space and time variables. In [22], they consider a two-parameter singularly perturbed parabolic problem with time-delay. They discretized the problem using an exponentially fitted scheme in the spatial direction and the Crank-Nicolson method in the time direction on a uniform mesh. In [3], the authors employ a fitted cubic spline scheme to solve a two-parameter singularly perturbed time-delayed problem. This scheme combines the θ-method on a uniform mesh in the time direction and a cubic spline scheme on a uniform mesh in the spatial direction. In [1], the authors focus on a two-parameter singularly perturbed system of partial differential equations with discontinuous coefficients. This discontinuity and small values of the perturbation parameters cause interior and boundary layers to appear in the solution. They employ a central finite-difference approach on a piecewise uniform Shishkin mesh in the spatial direction and an implicit Euler scheme on a uniform mesh in the time direction. In [29], the authors used a Hermite wavelet-based numerical method to solve two parameters singularly perturbed non-linear Benjamina-Bona-Mohany equation. The method involves time discretisation of Hermite wavelet series approximations with collocation technique. In [32], the authors used a finite element method on a Bakhvalov-type mesh to solve a two-parameter singularly perturbed two-point boundary value problem. Utilising individual interpolation, they obtain the errors for Lagrange interpolation and then prove the optimal order of convergence. In [2], the authors consider a singularly perturbed initial-boundary value problem with two parameters. They propose a fully-discrete numerical method combining the Crank-Nicolson scheme for time variables and the streamline-diffusion finite element method for the spatial variable. In [30], the authors consider a two-dimensional singularly perturbed convection-reaction-diffusion elliptic type problem with two parameters. Furthermore, the authors assume that jump discontinuities exist in the source term along the x-axis and y-axis. They employ an upwind finite-difference technique with an appropriate layer-adapted piecewise uniform Shishkin mesh to approximate the numerical solution.

In [26], the author proposed a numerical method based on an upwind finite difference operator to solve a two-parameter problem. They proved that the method is first-order convergent in space and time. The paper [15] presents a higher-order uniformly convergent method based on finite elements for a two-parameter parabolic singular perturbation problem. In [13], researchers have developed a hybrid scheme for one-dimensional singularly perturbed parabolic problems with two small parameters. The method presented is a composition of an upwind, midpoint upwind and central difference operator on a piecewise uniform Shishkin mesh. In contrast, the implicit Euler method is used for time stepping on a uniform mesh. A numerical approximation of a two-parameter parabolic problem with discontinuous data is proposed in [6]. In [11], the authors present an almost first-order convergent scheme for a two-parameter singularly perturbed problem with a time-delay. The method combines an upwind scheme for spatial discretisation and the implicit Euler

scheme for time discretisation. A hybrid difference approach for time dependent problems is studied in [14,27,31]. The author established almost second-order convergence in space direction and first-order convergence in time.

Due to their large applications, it is essential to develop robust numerical methods for solving two-parameter singular perturbation problems. This field of research is still developing, and this paper is a step forward in developing a parameter-uniform numerical method to solve a two-parameter singular perturbation problem with a large spatial delay or shift. Moreover, the paper presents rigorous consistency, stability and convergence analysis of the proposed method and illustrates numerical results to support theoretical estimates.

2. Continuous problem

Let $D = \Omega \times \Delta := (0,2) \times (0,T]$ and consider the following two-parameter, non-homogeneous initial boundary value problem

$$\left.\begin{aligned} &\mathcal{L}u(x,t) = \varepsilon u_{xx}(x,t) + \mu p(x,t)u_x(x,t) - q(x,t)u(x,t) + r(x,t)u(x-1,t) - u_t(x,t)\\ &\quad = g(x,t) \text{ in } D,\\ &u(x,t) = \psi_1(x,t) \text{ in } \Gamma_1 := \{(x,t),\ x \in [-1,0],\ t \in [0,T]\},\\ &u(x,t) = \psi_2(x,t) \text{ on } \Gamma_2 := \{(x,0),\ x \in [0,2]\},\\ &u(x,t) = \psi_3(x,t) \text{ on } \Gamma_3 := \{(2,t),\ t \in [0,T]\}, \end{aligned}\right\} \tag{1}$$

where $0 < \varepsilon < 1$ and $0 < \mu < 1$ are small parameters, $g(x,t)$, $p(x,t)$, $q(x,t)$ and $r(x,t)$ are sufficiently smooth functions such that

$$p(x,t) \geq p_0 > 0, \quad q(x,t) \geq q_0 > 0, \quad r(x,t) \geq r_0 > 0, \quad (q-r)(x,t) \geq \kappa > 0. \tag{2}$$

Moreover, we assume that the initial-boundary data ψ_1, ψ_2 and ψ_3 are smooth functions on their respective domain and satisfy the compatibility conditions given below

$$\begin{cases} \psi_2(0,0) = \psi_1(0,0), \quad \psi_2(2,0) = \psi_3(2,0),\\ \varepsilon\dfrac{\partial^2\psi_2(0,0)}{\partial x^2} + \mu p(0,0)\dfrac{\partial\psi_2(0,0)}{\partial x} - q(0,0)\psi_2(0,0)\\ \quad + r(0,0)\psi_1(-1,0) - \dfrac{\partial\psi_1(0,0)}{\partial t} = g(0,0),\\ \varepsilon\dfrac{\partial^2\psi_2(2,0)}{\partial x^2} + \mu p(2,0)\dfrac{\partial\psi_2(2,0)}{\partial x} - q(2,0)\psi_2(2,0)\\ \quad + r(2,0)\psi_2(1,0) - \dfrac{\partial\psi_3(2,0)}{\partial t} = g(2,0). \end{cases} \tag{3}$$

Let $D_1 = \Omega_1 \times \Delta := (0,1) \times (0,T]$ and $D_2 = \Omega_2 \times \Delta := [1,2) \times (0,T]$. Rewriting (1) as

$$\mathcal{L}u(x,t) = \mathcal{G}(x,t),$$

where

$$\mathcal{L}u(x,t) = \begin{cases} \mathcal{L}_1 u = \varepsilon u_{xx}(x,t) + \mu p(x,t)u_x(x,t) - q(x,t)u(x,t) \\ \qquad\qquad - u_t(x,t) \quad \text{if } (x,t) \in D_1, \\ \mathcal{L}_2 u = \varepsilon u_{xx}(x,t) + \mu p(x,t)u_x(x,t) - q(x,t)u(x,t) + r(x,t)u(x-1,t) \\ \qquad\qquad - u_t(x,t) \quad \text{if } (x,t) \in D_2, \end{cases} \tag{4}$$

and

$$\mathcal{G}(x,t) = \begin{cases} g(x,t) - r(x,t)\psi_1(x-1,t) & \text{if } (x,t) \in D_1, \\ g(x,t) & \text{if } (x,t) \in D_2 \end{cases} \tag{5}$$

with

$$\begin{cases} u(0,t) = \psi_1(0,t), \quad u(1^-,t) = u(1^+,t), \quad u_x(1^-,t) = u_x(1^+,t), \\ u(2,t) = \psi_3(2,t), \quad u(x,0) = \psi_2(x,0). \end{cases} \tag{6}$$

Under these assumptions, the solution of (1) exists and is unique [19]. The solution exhibits twin boundary layers due to the presence of perturbation parameters and an interior layer due to the presence of the shift [26]. Let $\gamma = \min\limits_{(x,t)\in\bar{D}} \left(\frac{q(x,t)-r(x,t)}{p(x,t)}\right)$. If $\frac{\mu^2}{\varepsilon} \leq \frac{\gamma}{p_0}$, the boundary layers of width $\mathcal{O}(\sqrt{\varepsilon})$ appear near the boundaries $x = 0$ and $x = 2$, as well as on the left and right neighbourhoods of $x = 1$ due to the large shift. If $\frac{\mu^2}{\varepsilon} \geq \frac{\gamma}{p_0}$, the boundary layers of width $\mathcal{O}(\mu)$ appear in a left neighbourhood of $x = 1$ and $x = 2$, while a boundary layer of width $\mathcal{O}\left(\frac{\varepsilon}{\mu}\right)$ appears in a right neighbourhood of $x = 0$ and $x = 1$.

Lemma 2.1: *Suppose* $\mathcal{P}(x,t) \in C^0(\bar{D}) \cap C^2(D)$ *satisfies* $\mathcal{P}(0,t) \geq 0, \mathcal{P}(2,t) \geq 0, \mathcal{P}(x,0) \geq 0$ *with* $\mathcal{L}_1\mathcal{P}(x,t) \leq 0$ *for all* $(x,t) \in (0,1) \times [0,T]$ *and* $\mathcal{L}_2\mathcal{P}(x,t) \leq 0$ *for all* $(x,t) \in [1,2) \times [0,T]$. *Then* $\mathcal{P}(x,t) \geq 0$ *for all* $(x,t) \in \bar{D}$.

Proof: Choose $(x^k,t^k) \in \bar{D}$ such that $\mathcal{P}(x^k,t^k) = \min\limits_{(x,t)\in\bar{D}} \mathcal{P}(x,t)$. Consequently,

$$\mathcal{P}_x(x^k,t^k) = 0, \quad \mathcal{P}_t(x^k,t^k) = 0 \quad \text{and} \quad \mathcal{P}_{xx}(x^k,t^k) > 0. \tag{7}$$

Suppose $\mathcal{P}(x^k,t^k) < 0$ and it follows that $(x^k,t^k) \notin \Gamma^* := \Gamma_1^* \cup \Gamma_2 \cup \Gamma_3$, where $\Gamma_1^* := \{(0,t),\ t \in [0,T]\}$.

Case I: If $(x^k,t^k) \in (0,1) \times [0,T]$, then

$$\begin{aligned} \mathcal{L}_1\mathcal{P}(x^k,t^k) &= \varepsilon\mathcal{P}_{xx}(x^k,t^k) + \mu p(x^k,t^k)\mathcal{P}_x(x^k,t^k) - q(x^k,t^k)\mathcal{P}(x^k,t^k) - \mathcal{P}_t(x^k,t^k) \\ &> 0, \text{ from (2) and (7).} \end{aligned}$$

Case II: If $(x^k, t^k) \in [1,2) \times [0,T]$, then

$$\begin{aligned}\mathcal{L}_2\mathcal{P}(x^k,t^k) &= \varepsilon\mathcal{P}_{xx}(x^k,t^k) + \mu p(x^k,t^k)\mathcal{P}_x(x^k,t^k) - q(x^k,t^k)\mathcal{P}(x^k,t^k)\\ &\quad + r(x^k,t^k)\mathcal{P}(x^k-1,t^k) - \mathcal{P}_t(x^k,t^k)\\ &\geq \varepsilon\mathcal{P}_{xx}(x^k,t^k) - (q-r)(x^k,t^k)\mathcal{P}(x^k,t^k)\\ &> 0, \text{ from (2) and (7).}\end{aligned}$$

A contradiction to the assumption that $\mathcal{L}_1\mathcal{P} \leq 0$ for all $(x,t) \in (0,1) \times [0,T]$ and $\mathcal{L}_2\mathcal{P} \leq 0$ for all $(x,t) \in [1,2) \times [0,T]$. Consequently, the required result follows from a contradiction. ■

As a consequence of Lemma 2.1, obtaining the following stability estimate is straightforward.

Lemma 2.2: *Let u be the solution of (1). Then* $\|u\|_{\bar{D}} \leq \|u\|_{\Gamma^*} + \frac{1}{\kappa}\|\mathcal{G}\|_{\bar{D}}$.

Proof: Define $\theta_\pm(x,t) = \|u\|_{\Gamma^*} + \dfrac{1}{\kappa}\|\mathcal{G}\|_{\bar{D}} \pm u(x,t)$, $(x,t) \in \bar{D}$. Then $\theta_\pm(x,t) \geq 0$ for all $(x,t) \in \Gamma^*$ and if $(x,t) \in (0,1) \times [0,T]$, it follows that

$$\begin{aligned}\mathcal{L}_1\theta_\pm(x,t) &= -q(x,t)\left(\|u\|_{\Gamma^*} + \frac{\|\mathcal{G}\|_{\bar{D}}}{\kappa}\right) \pm \mathcal{L}_1 u(x,t)\\ &= -q(x,t)\left(\|u\|_{\Gamma^*} + \frac{\|\mathcal{G}\|_{\bar{D}}}{\kappa}\right) \pm \mathcal{G}(x,t) \leq 0\end{aligned}$$

Similarly, for $(x,t) \in [1,2) \times [0,T]$

$$\mathcal{L}_2\theta_\pm(x,t) = -(q-r)(x,t)\left(\|u\|_{\Gamma^*} + \frac{\|\mathcal{G}\|_{\bar{D}}}{\kappa}\right) \pm \mathcal{L}_2 u(x,t) \leq 0.$$

The required result thus follows from Lemma 2.1. ■

Lemma 2.3: *Let u be the solution of (1). Then, for* $1 \leq k + 2m \leq 3$

$$\left\|\frac{\partial^{k+m}u}{\partial x^k \partial t^m}\right\|_{\bar{D}} \leq \begin{cases}\dfrac{C}{(\sqrt{\varepsilon})^k} & \text{if } p_0\mu^2 \leq \gamma\,\varepsilon,\\ C\left(\dfrac{\mu}{\varepsilon}\right)^k\left(\dfrac{\mu^2}{\varepsilon}\right)^m & \text{if } p_0\mu^2 \geq \gamma\,\varepsilon,\end{cases}$$

where C is a constant independent of ε *and* μ.

Proof: We obtain the bounds of the solution and its derivatives by splitting the argument into two cases $p_0\mu^2 \leq \gamma\,\varepsilon$ and $p_0\mu^2 \geq \gamma\,\varepsilon$.

Case I: Let $p_0\mu^2 \le \gamma\,\varepsilon$. Consider the stretching variables $\tilde{x} = \dfrac{x}{\sqrt{\varepsilon}}$, $\tilde{t} = t$ to obtain the transformed problem on the domain $\tilde{D} = (0, 2/\sqrt{\varepsilon}) \times (0, T]$:

$$\begin{cases} \left(\tilde{u}_{\tilde{x}\tilde{x}} + \frac{\mu}{\sqrt{\varepsilon}}\tilde{p}\tilde{u}_{\tilde{x}} - \tilde{q}\tilde{u} - \tilde{u}_{\tilde{t}}\right)(\tilde{x},\tilde{t}) + \tilde{r}(\tilde{x},\tilde{t})\tilde{u}(\tilde{x}-1,\tilde{t}) = \tilde{g}(\tilde{x},\tilde{t}) \text{ in } \tilde{D}, \\ \tilde{u}(\tilde{x},\tilde{t}) = \tilde{\psi_1}(\tilde{x},\tilde{t}) \text{ on } \tilde{\Gamma_1}, \\ \tilde{u}(\tilde{x},\tilde{t}) = \tilde{\psi_2}(\tilde{x},\tilde{t}) \text{ on } \tilde{\Gamma_2}, \\ \tilde{u}(\tilde{x},\tilde{t}) = \tilde{\psi_3}(\tilde{x},\tilde{t}) \text{ on } \tilde{\Gamma_3}. \end{cases}$$

Using the condition $p_0\mu^2 \le \gamma\,\varepsilon$ and result from [19], we have

$$\left\| \frac{\partial^{k+m}\tilde{u}}{\partial\tilde{x}^k\partial\tilde{t}^m} \right\|_{N_{\lambda,\xi}} \le C(1 + \|\tilde{u}\|_{\tilde{D}}), \quad 1 \le k + 2m \le 3,$$

where $N_{\lambda,\xi}$ is the rectangle $(\lambda - \xi, \lambda + \xi) \times (0, T] \cap \tilde{D}$ for any $\lambda \in (0, 2/\sqrt{\varepsilon})$ and $\xi > 0$. Now, we return back to the original variable to get

$$\begin{aligned} \left\| \frac{\partial^{k+m}u}{\partial x^k\partial t^m} \right\|_{\bar{D}} &\le C\varepsilon^{\frac{-k}{2}}(1 + \|u\|_{\bar{D}}), \quad 1 \le k + 2m \le 3 \\ &\le \frac{C}{(\sqrt{\varepsilon})^k}, \text{ from Lemma 2.2.} \end{aligned}$$

Case II: Let $p_0\mu^2 \ge \gamma\,\varepsilon$. Consider the stretching variables $\tilde{x} = \dfrac{\mu x}{\varepsilon}$, $\tilde{t} = \dfrac{\mu^2 t}{\varepsilon}$ to obtain the following transformed problem on the domain $\tilde{D} = (0, 2\mu/\varepsilon) \times (0, \mu^2 T/\varepsilon]$:

$$\begin{cases} \left(\tilde{u}_{\tilde{x}\tilde{x}} + \tilde{p}\tilde{u}_{\tilde{x}} - \dfrac{\tilde{q}\varepsilon}{\mu^2}\tilde{u} - \tilde{u}_{\tilde{t}}\right)(\tilde{x},\tilde{t}) + \dfrac{\tilde{r}\varepsilon}{\mu^2}(\tilde{x},\tilde{t})\tilde{u}(\tilde{x}-1,\tilde{t}) = \dfrac{\varepsilon}{\mu^2}\tilde{g}(\tilde{x},\tilde{t}) \text{ in } \tilde{D}, \\ \tilde{u}(\tilde{x},\tilde{t}) = \tilde{\psi_1}(\tilde{x},\tilde{t}) \text{ on } \tilde{\Gamma_1}, \\ \tilde{u}(\tilde{x},\tilde{t}) = \tilde{\psi_2}(\tilde{x},\tilde{t}) \text{ on } \tilde{\Gamma_2}, \\ \tilde{u}(\tilde{x},\tilde{t}) = \tilde{\psi_3}(\tilde{x},\tilde{t}) \text{ on } \tilde{\Gamma_3}. \end{cases}$$

Repeating the argument given in Case I, we get

$$\left\| \frac{\partial^{k+m}u}{\partial x^k\partial t^m} \right\|_{\bar{D}} \le C\left(\frac{\mu}{\varepsilon}\right)^k\left(\frac{\mu^2}{\varepsilon}\right)^m, \quad 1 \le k + 2m \le 3. \qquad \blacksquare$$

Corollary 2.4: *Let $u(x,t)$ be the solution of* (1) *and Lemmas 2.1 and 2.2 hold. Then*

$$\|u_{tt}\|_{\bar{D}} \le \begin{cases} C & \text{if } p_0\mu^2 \le \gamma\,\varepsilon, \\ C\mu^4\varepsilon^{-2} & \text{if } p_0\mu^2 \ge \gamma\,\varepsilon. \end{cases}$$

3. Time discretization

Let $\mathbb{T}_t^M = \{t_k = kT/M, k = 0, \ldots, M\}$ be an equidistant mesh in the time direction. We use the Crank-Nicolson method to discretize the problem (1) in the time variable. The semidiscrete problem on $\mathbb{T}_t^M$ thus reads

$$\left.\begin{aligned}&\left(\varepsilon U_{xx} + \mu p U_x - lU\right)(x, t_{k+1}) + r(x, t_{k+1})U(x-1, t_{k+1})\\&\quad = \left(-\varepsilon U_{xx} - \mu p U_x + mU\right)(x, t_k) - r(x, t_k)U(x-1, t_k) + g(x, t_{k+1}) + g(x, t_k),\\&\quad x \in \Omega,\ k = 0, 1, \ldots, M-1\end{aligned}\right\} \tag{8}$$

such that

$$\left.\begin{aligned}&U(x, t_{k+1}) = \psi_1(x, t_{k+1}), \quad -1 \le x \le 0, \quad -1 \le k \le M-1,\\&U(x, 0) = \psi_2(x, 0), \quad 0 \le x \le 2,\\&U(2, t_{k+1}) = \psi_3(2, t_{k+1}), \quad -1 \le k \le M-1,\end{aligned}\right\} \tag{9}$$

where $l(x, t_{k+1}) = q(x, t_{k+1}) + \dfrac{2}{\Delta t}$, $m(x, t_k) = q(x, t_k) - \dfrac{2}{\Delta t}$ and $U(x, t_{k+1})$ is the numerical approximation of the continuous solution $u(x, t)$ at $(k+1)$th time step.

Let us rewrite (8) as

$$\tilde{\mathcal{L}}_n U(x, t_{k+1}) = \tilde{\mathcal{G}}_n(x, t_{k+1}), \quad n = 1, 2, \tag{10}$$

where

$$\begin{aligned}\tilde{\mathcal{L}}_1 U(x, t_{k+1}) &= \left(\varepsilon U_{xx} + \mu p U_x - lU\right)(x, t_{k+1}) \quad \text{if } x \in (0, 1),\\ \tilde{\mathcal{L}}_2 U(x, t_{k+1}) &= \left(\varepsilon U_{xx} + \mu p U_x - lU\right)(x, t_{k+1}) + r(x, t_{k+1})U(x-1, t_{k+1}) \quad \text{if } x \in [1, 2),\end{aligned}$$

and

$$\begin{aligned}\tilde{\mathcal{G}}_1(x, t_{k+1}) &= \left(-\varepsilon U_{xx} - \mu p U_x + mU\right)(x, t_k) - r(x, t_{k+1})\psi_1(x-1, t_{k+1})\\&\quad - r(x, t_k))\psi_1(x-1, t_k) + g(x, t_{k+1}) + g(x, t_k) \quad \text{if } x \in (0, 1),\\ \tilde{\mathcal{G}}_2(x, t_{k+1}) &= \left(-\varepsilon U_{xx} - \mu p U_x + mU\right)(x, t_k) - r(x, t_k))\psi_1(x-1, t_k)\\&\quad + g(x, t_{k+1}) + g(x, t_k) \quad \text{if } x \in [1, 2).\end{aligned}$$

The operator $\tilde{\mathcal{L}}_n$ satisfies the following minimum principle.

Lemma 3.1: *Let $\chi(x, t_{k+1})$ be a smooth function such that $\chi(x, t_{k+1}) \ge 0$ for $x = 0, 2$ with $\tilde{\mathcal{L}}_1\chi(x, t_{k+1}) \le 0$ for all $x \in (0, 1)$ and $\tilde{\mathcal{L}}_2\chi(x, t_{k+1}) \le 0$ for all $x \in [1, 2)$. Then $\chi(x, t_{k+1}) \ge 0$ for all $x \in \bar{\Omega}$.*

Proof: Let $\chi(\alpha, t_{k+1}) = \min_{x\in\bar{\Omega}} \chi(x, t_{k+1})$ for some $\alpha \in \bar{\Omega}$. Then

$$\chi_x(\alpha, t_{k+1}) = 0 \quad \text{and} \quad \chi_{xx}(\alpha, t_{k+1}) > 0. \tag{11}$$

Suppose $\chi(\alpha, t_{k+1}) < 0$, therefore $\alpha \notin \{0, 2\}$ because $\chi(x, t_{k+1}) \ge 0$ for $x = 0, 2$.

Case I: If $\alpha \in (0,1)$

$$\begin{aligned}\tilde{\mathcal{L}}_1\chi(\alpha,t_{k+1}) &= \left(\varepsilon\chi_{xx}+\mu p\chi_x - l\chi\right)(\alpha,t_{k+1})\\ &> 0, \quad \text{from (2) and (11).}\end{aligned}$$

Case II: If $\alpha \in [1,2)$

$$\begin{aligned}\tilde{\mathcal{L}}_2\chi(\alpha,t_{k+1}) &= \left(\varepsilon\chi_{xx}+\mu p\chi_x - l\chi\right)(\alpha,t_{k+1}) + r(\alpha,t_{k+1})\chi(\alpha-1,t_{k+1})\\ &\geq \varepsilon\chi_{xx}(\alpha,t_{k+1}) - l(\alpha,t_{k+1})\chi(\alpha,t_{k+1}) + r(\alpha,t_{k+1})\chi(\alpha,t_{k+1})\\ &\geq \varepsilon\chi_{xx}(\alpha,t_{k+1}) - \left(q(\alpha,t_{k+1}) + \frac{2}{\Delta t} - r(\alpha,t_{k+1})\right)\chi(\alpha,t_{k+1})\\ &= \varepsilon\chi_{xx}(\alpha,t_{k+1}) - (q-r)(\alpha,t_{k+1})\chi(\alpha,t_{k+1}) - \frac{2}{\Delta t}\chi(\alpha,t_{k+1})\\ &> 0, \quad \text{from (2) and (11).}\end{aligned}$$

It contradicts the assumption, and hence the result follows from a contradiction. ■

Lemma 3.2: *Let $U(x,t_{k+1})$ be the solution of* (10)*. Then*

$$\|U(x,t_{k+1})\|_{\bar{\Omega}} \leq \max\left\{|U(0,t_{k+1})|, |U(2,t_{k+1})|, \frac{1}{\kappa}\|\tilde{\mathcal{G}}\|_{\bar{\Omega}}\right\} \quad \text{for all } x \in [0,2].$$

Proof: Consider $M_{\pm}(x,t_{k+1}) = \max\{|U(0,t_{k+1})|, \frac{1}{\kappa}\|\tilde{\mathcal{G}}_1\|_{\bar{\Omega}_1}\} \pm U(x,t_{k+1})$ for all $x \in (0,1)$. Then, $M_{\pm}(0,t_{k+1}) \geq 0$. Moreover, for $x \in (0,1)$

$$\begin{aligned}\tilde{\mathcal{L}}_1 M_{\pm}(x,t_{k+1}) &= -l(x,t_{k+1})\max\left\{|U(0,t_{k+1})|, \frac{1}{\kappa}\|\tilde{\mathcal{G}}_1\|_{\bar{\Omega}_1}\right\} \pm \tilde{\mathcal{L}}_1 U(x,t_{k+1})\\ &= -l(x,t_{k+1})\max\left\{|U(0,t_{k+1})|, \frac{1}{\kappa}\|\tilde{\mathcal{G}}_1\|_{\bar{\Omega}_1}\right\} \pm \tilde{\mathcal{G}}_1(x,t_{k+1})\\ &\leq 0, \quad \text{from (2).}\end{aligned}$$

Similarly, consider $M_{\pm}(x,t_{k+1}) = \max\{|U(2,t_{k+1})|, \frac{1}{\kappa}\|\tilde{\mathcal{G}}_2\|_{\bar{\Omega}_2}\} \pm U(x,t_{k+1})$ for all $x \in [1,2)$. Then, $M_{\pm}(2,t_{k+1}) \geq 0$ and for $x \in [1,2)$ we compute

$$\begin{aligned}&\tilde{\mathcal{L}}_2 M_{\pm}(x,t_{k+1})\\ &= -(l-r)(x,t_{k+1})\max\left\{|U(2,t_{k+1})|, \frac{1}{\kappa}\|\tilde{\mathcal{G}}_2\|_{\bar{\Omega}_2}\right\} \pm \tilde{\mathcal{L}}_2 U(x,t_{k+1})\\ &= -\left(q(x,t_{k+1}) + \frac{2}{\Delta t} - r(x,t_{k+1})\right)\max\left\{|U(2,t_{k+1})|, \frac{1}{\kappa}\|\tilde{\mathcal{G}}_2\|_{\bar{\Omega}_2}\right\} \pm \tilde{\mathcal{G}}_2(x,t_{k+1})\\ &\leq 0, \quad \text{from (2).}\end{aligned}$$

Consequently, from Lemma 3.1 the required result follows. ■

The local truncation error of the semidiscretized problem (10) is given by $e_{k+1} := \hat{U}(x, t_{k+1}) - u(x, t_{k+1})$, where $\hat{U}$ is the solution of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_n \hat{U}(x, t_{k+1}) = \hat{\mathcal{G}}(x, t_{k+1}), \quad x \in \Omega, \quad 0 \le k \le M-1, \\ &\hat{U}(x, t_{k+1}) = \psi_1(x, t_{k+1}), \quad -1 \le x \le 0, \quad -1 \le k \le M-1, \\ &\hat{U}(x, 0) = \psi_2(x, 0), \quad x \in \bar{\Omega}, \\ &\hat{U}(2, t_{k+1}) = \psi_3(2, t_{k+1}), \quad -1 \le k \le M-1, \end{aligned}\right\} \tag{12}$$

where $\tilde{\mathcal{L}}_n$ is defined in (10), and

$$\hat{\mathcal{G}} = \begin{cases} \left(-\varepsilon u_{xx} - \mu p u_x + mu\right)(x, t_k) - r(x, t_{k+1})\psi_1(x-1, t_{k+1}) \\ \quad -r(x, t_k)\psi_1(x-1, t_k) + g(x, t_{k+1}) + g(x, t_k) & \text{if } x \in (0, 1), \\ \left(-\varepsilon u_{xx} - \mu p u_x + mu\right)(x, t_k) - r(x, t_k))\psi_1(x-1, t_k) \\ \quad +g(x, t_{k+1}) + g(x, t_k) & \text{if } x \in [1, 2). \end{cases}$$

Lemma 3.3: *For some constant C, the local truncation error at* $(k+1)$*th time step satisfies the following bound*

$$\left\| e_{k+1} \right\|_\infty \le C(\Delta t)^3, \quad -1 \le k \le M-1.$$

Proof: For proof, see [9]. ■

The global error $E_k := u(x, t_k) - U(x, t_k)$ of the semidiscretized problem is the contribution of the local truncation error at each time step. Then, it follows that

$$\left\| E_{k+1} \right\|_\infty = \left\| \sum_{i=1}^{k} \hat{e}_i \right\|_\infty \le \left\| \hat{e}_1 \right\|_\infty + \left\| \hat{e}_2 \right\|_\infty + \ldots + \left\| \hat{e}_k \right\|_\infty \le C\Delta t^2. \tag{13}$$

This in turn ensures the uniform convergence of the time semidiscretization process.

Lemma 3.4: *Let* $U(x, t_{k+1})$ *be the solution of* (10). *Then*

$$\left\| \frac{d^k U}{dx^k} \right\| \le \frac{C}{(\sqrt{\varepsilon})^k} \left(1 + \left(\frac{\mu}{\sqrt{\varepsilon}}\right)^k\right) \max\left\{ \|U\|, \|\tilde{\mathcal{G}}\| \right\}, \quad k = 1, 2,$$

$$\left\| \frac{d^3 U}{dx^3} \right\| \le \frac{C}{(\sqrt{\varepsilon})^3} \left(1 + \left(\frac{\mu}{\sqrt{\varepsilon}}\right)^3\right) \max\left\{ \|U\|, \|\tilde{\mathcal{G}}\| \right\} + \frac{1}{\varepsilon} \left\| \frac{\mathrm{d}\tilde{\mathcal{G}}}{\mathrm{d}x} \right\|,$$

where C is the constant independent of μ *and* ε.

Proof: For proof, see [16]. ■

The solution $U(x, t_{k+1})$ of the semidiscretized problem (10) admit a decomposition into smooth and singular components [20]. We write

$$U(x, t_{k+1}) := X(x, t_{k+1}) + Z(x, t_{k+1}),$$

where

$$X(x, t_{k+1}) = \begin{cases} b(x, t_{k+1}), & x \in [0, 1), \\ c(x, t_{k+1}), & x \in [1, 2], \end{cases} \tag{14}$$

and

$$Z(x, t_{k+1}) = \begin{cases} z_1(x, t_{k+1}), & x \in [0, 1), \\ z_2(x, t_{k+1}), & x \in [1, 2]. \end{cases} \tag{15}$$

Here, the smooth component $b(x, t_{k+1})$ is the solution of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_1 b(x, t_{k+1}) = \tilde{\mathcal{G}}_1(x, t_{k+1}), \quad x \in (0, 1), \\ &b(0, t_{k+1}), \quad b(1, t_{k+1}) \text{ are to be chosen lately,} \end{aligned}\right\} \tag{16}$$

and $c(x, t_{k+1})$ is the solution of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_2 c(x, t_{k+1}) = \tilde{\mathcal{G}}_2(x, t_{k+1}), \quad x \in (1, 2), \\ &c(x, t_{k+1}) = b(x, t_{k+1}) \text{ on } [0, 1), \\ &c(1, t_{k+1}), \quad c(2, t_{k+1}) \text{ are to be chosen lately.} \end{aligned}\right\} \tag{17}$$

Moreover, the singular components $z_1(x, t_{k+1})$ and $z_2(x, t_{k+1})$ satisfy the following

$$\tilde{\mathcal{L}}_1 z_1(x, t_{k+1}) = 0, \quad x \in (0, 1), \quad \tilde{\mathcal{L}}_2 z_2(x, t_{k+1}) = 0, \quad x \in (1, 2), \tag{18}$$

$$z_1(0, t_{k+1}) = U(0, t_{k+1}) - b(0, t_{k+1}),$$

$$z_2(1, t_{k+1}) - z_1(1, t_{k+1}) = b(1, t_{k+1}) - c(1, t_{k+1}), \tag{19}$$

$$\frac{\mathrm{d}z_2}{\mathrm{d}x}(1, t_{k+1}) - \frac{\mathrm{d}z_1}{\mathrm{d}x}(1, t_{k+1}) = \frac{\mathrm{d}b}{\mathrm{d}x}(1, t_{k+1}) - \frac{\mathrm{d}c}{\mathrm{d}x}(1, t_{k+1}), \tag{20}$$

$$z_2(2, t_{k+1}) = U(2, t_{k+1}) - c(2, t_{k+1}).$$

The next lemmas provide bounds on the derivatives of $X(x, t_{k+1})$ and $Z(x, t_{k+1})$ with respect to x. We derive bounds for these components separately for $p_0\mu^2 \le \gamma\,\varepsilon$ and $p_0\mu^2 \ge \gamma\,\varepsilon$.

3.1. Estimates for the smooth components when $p_0\mu^2 \le \gamma\,\varepsilon$

Lemma 3.5: *Let $b(x, t_{k+1})$ and $c(x, t_{k+1})$ be the solutions of* (16) *and* (17)*, respectively. Then for $p_0\mu^2 \le \gamma\,\varepsilon$ and $k = 0, 1, 2, 3$*

$$\|b^{(k)}\|_{\bar{\Omega}_1} \le C, \quad \|c^{(k)}\|_{\bar{\Omega}_2} \le C.$$

Proof: Decompose the smooth components $b(x,t_{k+1})$ and $c(x,t_{k+1})$ as

$$\begin{aligned} b(x,t_{k+1}) &= b_0(x,t_{k+1}) + \sqrt{\varepsilon}b_1(x,t_{k+1}) + (\sqrt{\varepsilon})^2 b_2(x,t_{k+1}) \\ &\quad + (\sqrt{\varepsilon})^3 b_3(x,t_{k+1}), \quad x \in \bar{\Omega}_1, \\ c(x,t_{k+1}) &= c_0(x,t_{k+1}) + \sqrt{\varepsilon}c_1(x,t_{k+1}) + (\sqrt{\varepsilon})^2 c_2(x,t_{k+1}) \\ &\quad + (\sqrt{\varepsilon})^3 c_3(x,t_{k+1}), \quad x \in \bar{\Omega}_2, \end{aligned}$$

where the functions b_0 and c_0 are the solutions of

$$\left.\begin{aligned} -(lb_0)(x,t_{k+1}) &= (mb_0)(x,t_k) - r(x,t_{k+1})\psi_1(x-1,t_{k+1}) - r(x,t_k)\psi_1(x-1,t_k) \\ &\quad + g(x,t_k) + g(x,t_{k+1}), \quad x \in \bar{\Omega}_1, \\ -(lc_0)(x,t_{k+1}) + r(x,t_{k+1})b_0(x-1,t_{k+1}) &= (mc_0)(x,t_k) - r(x,t_k)b_0(x-1,t_k) \\ &\quad + g(x,t_k) + g(x,t_{k+1}), \quad x \in \bar{\Omega}_2, \end{aligned}\right\}$$

the functions b_1 and c_1 are the solutions of

$$\left.\begin{aligned} (lb_1)(x,t_{k+1}) &= \left(\sqrt{\varepsilon}b_0'' + \frac{\mu}{\sqrt{\varepsilon}}pb_0'\right)(x,t_{k+1}) \\ &\quad + \left(\sqrt{\varepsilon}b_0'' + \frac{\mu}{\sqrt{\varepsilon}}pb_0' - mb_1\right)(x,t_k), \quad x \in \bar{\Omega}_1, \\ (lc_1)(x,t_{k+1}) - r(x,t_{k+1})b_1(x-1,t_{k+1}) &= \left(\sqrt{\varepsilon}c_0'' + \frac{\mu}{\sqrt{\varepsilon}}pc_0'\right)(x,t_{k+1}) \\ &\quad + r(x,t_k)b_1(x-1,t_k) + \left(\sqrt{\varepsilon}c_0'' + \tfrac{\mu}{\sqrt{\varepsilon}}pc_0' - mc_1\right)(x,t_k), \quad x \in \bar{\Omega}_2, \end{aligned}\right\}$$

the functions b_2 and c_2 are the solutions of

$$\left.\begin{aligned} (lb_2)(x,t_{k+1}) &= \left(\sqrt{\varepsilon}b_1'' + \frac{\mu}{\sqrt{\varepsilon}}pb_1'\right)(x,t_{k+1}) \\ &\quad + \left(\sqrt{\varepsilon}b_1'' + \frac{\mu}{\sqrt{\varepsilon}}pb_1' - mb_2\right)(x,t_k), \quad x \in \bar{\Omega}_1, \\ &(lc_2)(x,t_{k+1}) - r(x,t_{k+1})b_2(x-1,t_{k+1}) \\ &\quad = \left(\sqrt{\varepsilon}c_1'' + \frac{\mu}{\sqrt{\varepsilon}}pc_1'\right)(x,t_{k+1}) + r(x,t_k)b_2(x-1,t_k) \\ &\qquad + \left(\sqrt{\varepsilon}c_1'' + \tfrac{\mu}{\sqrt{\varepsilon}}pc_1' - mc_2\right)(x,t_k), \quad x \in \bar{\Omega}_2, \end{aligned}\right\}$$

and lastly, the functions b_3 and c_3 are the solutions of

$$\left.\begin{aligned} \tilde{\mathcal{L}}_1 b_3(x,t_{k+1}) &= \left(-\sqrt{\varepsilon}b_2'' - \frac{\mu}{\sqrt{\varepsilon}}pb_2'\right)(x,t_{k+1}) + \left(-\sqrt{\varepsilon}b_2'' - \frac{\mu}{\sqrt{\varepsilon}}pb_2'\right)(x,t_k) \\ &\quad - (\varepsilon b_3'' + \mu p b_3' - mb_3)(x,t_k), \quad x \in \Omega_1, \\ b_3(0,t_{k+1}) &= 0, \quad b_3(1,t_{k+1}) = 0, \end{aligned}\right\} \tag{22}$$

and

$$\left.\begin{aligned}\tilde{\mathcal{L}}_2 c_3(x,t_{k+1}) &= \left(-\sqrt{\varepsilon}c_2'' - \frac{\mu}{\sqrt{\varepsilon}}pc_2'\right)(x,t_{k+1}) - \left(\sqrt{\varepsilon}c_2'' + \frac{\mu}{\sqrt{\varepsilon}}pc_2'\right)(x,t_k)\\ &-(\varepsilon c_3'' + \mu p c_3' - m c_3)(x,t_k) - r(x,t_k)b_3(x-1,t_k), \quad x \in \Omega_2,\\ c_3(1,t_{k+1}) &= 0, \quad c_3(2,t_{k+1}) = 0, \quad c_3(x,t_{k+1}) = b_3(x,t_{k+1}) \text{ on } [0,1).\end{aligned}\right\} \tag{23}$$

Since, the functions $p(x,t_{k+1})$, $r(x,t_{k+1})$, $m(x,t_{k+1})$, $l(x,t_{k+1})$, $\psi_1(x,t_{k+1})$ and $g(x,t_{k+1})$ are sufficiently smooth and $p_0\mu^2 \le \gamma\,\varepsilon$, we have

$$\begin{aligned}\|b_0^{(k)}\|_{\bar{\Omega}_1} \le C, \quad \|c_0^{(k)}\|_{\bar{\Omega}_2} \le C, \quad &\text{for } 0 \le k \le 7,\\ \|b_1^{(k)}\|_{\bar{\Omega}_1} \le C, \quad \|c_1^{(k)}\|_{\bar{\Omega}_2} \le C, \quad &\text{for } 0 \le k \le 5,\\ \|b_2^{(k)}\|_{\bar{\Omega}_1} \le C, \quad \|c_2^{(k)}\|_{\bar{\Omega}_2} \le C, \quad &\text{for } 0 \le k \le 3.\end{aligned}$$

Now, using Lemma 2 from [12] to obtain $\|b_3\|_{\bar{\Omega}_1} \le C$ and $\|c_3\|_{\bar{\Omega}_2} \le C$. Since $p_0\mu^2 \le \gamma\,\varepsilon$, Lemma 3.4 asserts that $\|b_3'\|_{\bar{\Omega}_1} \le \frac{C}{\sqrt{\varepsilon}}$ and $\|c_3'\|_{\bar{\Omega}_2} \le \frac{C}{\sqrt{\varepsilon}}$. Moreover, from (22) and (23), we obtain

$$\|b_3^{(k)}\|_{\bar{\Omega}_1} \le \frac{C}{(\sqrt{\varepsilon})^k}, \quad \|c_3^{(k)}\|_{\bar{\Omega}_2} \le \frac{C}{(\sqrt{\varepsilon})^k}, \quad \text{for } k = 2,3.$$

Next, we choose $b(0,t_{k+1})$, $b(1,t_{k+1})$, $c(1,t_{k+1})$ and $c(2,t_{k+1})$ to be

$$\begin{aligned}b(0,t_{k+1}) &= (b_0 + \sqrt{\varepsilon}b_1 + \varepsilon b_2)(0,t_{k+1}), \quad b(1,t_{k+1}) = (b_0 + \sqrt{\varepsilon}b_1 + \varepsilon b_2)(1,t_{k+1}),\\ c(1,t_{k+1}) &= (c_0 + \sqrt{\varepsilon}c_1 + \varepsilon c_2)(1,t_{k+1}), \quad c(2,t_{k+1}) = (c_0 + \sqrt{\varepsilon}c_1 + \varepsilon c_2)(2,t_{k+1}).\end{aligned}$$

Since $b = b_0 + \sqrt{\varepsilon}b_1 + \varepsilon b_2 + \varepsilon^{3/2}b_3$, for $k = 0, 1, 2, 3$, we have

$$\|b^{(k)}\|_{\bar{\Omega}_1} \le \|b_0^{(k)}\|_{\bar{\Omega}_1} + \sqrt{\varepsilon}\|b_1^{(k)}\|_{\bar{\Omega}_1} + \varepsilon\|b_2^{(k)}\|_{\bar{\Omega}_1} + \varepsilon^{3/2}\|b_3^{(k)}\|_{\bar{\Omega}_1} \le C.$$

Similarly, we get $\|c^{(k)}\|_{\bar{\Omega}_2} \le C$, for $k = 0, 1, 2, 3$. ■

3.2. Estimates for the smooth components when $p_0\mu^2 \ge \gamma\varepsilon$

Similar to the previous case, here also, the decomposition of $b(x,t_{k+1})$ and $c(x,t_{k+1})$ is considered, but at two levels. We define operators $\tilde{\mathcal{L}}_1^*$, $\tilde{\mathcal{L}}_2^*$, $\tilde{\mathcal{L}}_3$, $\tilde{\mathcal{L}}_3^*$, $\tilde{\mathcal{L}}_4$, and $\tilde{\mathcal{L}}_4^*$ for this purpose. These operators are defined as:

$$\begin{aligned}&\tilde{\mathcal{L}}_1^* = \varepsilon D^2 + \mu p(x,t_k)D - m(x,t_k)I, \quad \tilde{\mathcal{L}}_2^* = \varepsilon D^2 + \mu p(x,t_k)D - m(x,t_k)I + r(x,t_k)E,\\ &\tilde{\mathcal{L}}_3 = \mu p(x,t_k)D - l(x,t_k)I, \quad \tilde{\mathcal{L}}_3^* = \mu p(x,t_k)D - m(x,t_k)I,\\ &\tilde{\mathcal{L}}_4 = \mu p(x,t_k)D - l(x,t_k)I + r(x,t_k)E \quad \text{and} \quad \tilde{\mathcal{L}}_4^* = \mu p(x,t_k)D - m(x,t_k)I + r(x,t_k)E.\end{aligned}$$

Here, $DU(x,t_k) = \dfrac{\mathrm{d}U}{\mathrm{d}x}(x,t_k)$ and $D^2U(x,t_k) = \dfrac{\mathrm{d}^2U}{\mathrm{d}x^2}(x,t_k)$ are the differential operators. Additionally, $IU(x,t_k) = U(x,t_k)$ is the identity operator and $EU(x,t_k) = EU(x-1,t_k)$ is

the shift operator. Finally, we consider the decomposition of $b(x, t_{k+1})$ and $c(x, t_{k+1})$ as

$$\begin{aligned} b(x, t_{k+1}; \varepsilon, \mu) &= b_0(x, t_{k+1}) + \varepsilon b_1(x, t_{k+1}; \mu) + \varepsilon^2 b_2(x, t_{k+1}; \mu) \\ &\quad + \varepsilon^3 b_3(x, t_{k+1}; \mu), \quad x \in \bar{\Omega}_1, \\ c(x, t_{k+1}; \varepsilon, \mu) &= c_0(x, t_{k+1}) + \varepsilon c_1(x, t_{k+1}; \mu) + \varepsilon^2 c_2(x, t_{k+1}; \mu) \\ &\quad + \varepsilon^3 c_3(x, t_{k+1}; \mu), \quad x \in \bar{\Omega}_2, \end{aligned}$$

where the functions b_0 and c_0 are the solutions of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_3 b_0(x, t_{k+1}) = -\tilde{\mathcal{L}}_3^* b_0(x, t_k) - r(x, t_{k+1})\psi_1(x-1, t_{k+1}) \\ &\quad -r(x, t_k)\psi_1(x-1, t_k) + g(x, t_k) + g(x, t_{k+1}), \quad x \in [0, 1), \\ &b_0(1, t_{k+1}) \text{ is to be chosen lately,} \end{aligned}\right\} \tag{24}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_4 c_0(x, t_{k+1}) = -\tilde{\mathcal{L}}_4^* c_0(x, t_k) + g(x, t_k) + g(x, t_{k+1}), \quad x \in [1, 2), \\ &c_0(2, t_{k+1}) \text{ is to be chosen lately,} \\ &c_0(x, t_{k+1}) = b_0(x, t_{k+1}) \text{ on } [0, 1), \end{aligned}\right\} \tag{25}$$

the functions b_1 and c_1 are the solutions of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_3 b_1(x, t_{k+1}) = -b_0''(x, t_{k+1}) - b_0''(x, t_k) - \tilde{\mathcal{L}}_3^* b_1(x, t_k), \quad x \in [0, 1), \\ &b_1(1, t_{k+1}) \text{ is to be chosen lately,} \end{aligned}\right\} \tag{26}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_4 c_1(x, t_{k+1}) = -c_0''(x, t_{k+1}) - c_0''(x, t_k) - \tilde{\mathcal{L}}_4^* c_1(x, t_k), \quad x \in [1, 2), \\ &c_1(2, t_{k+1}) \text{ is to be chosen lately,} \\ &c_1(x, t_{k+1}) = b_1(x, t_{k+1}) \text{ on } [0, 1), \end{aligned}\right\} \tag{27}$$

the functions b_2 and c_2 are the solutions of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_3 b_2(x, t_{k+1}) = -b_1''(x, t_{k+1}) - b_1''(x, t_k) - \tilde{\mathcal{L}}_3^* b_2(x, t_k), \quad x \in [0, 1), \\ &b_2(1, t_{k+1}) = 0, \end{aligned}\right\} \tag{28}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_4 c_2(x, t_{k+1}) = -c_1''(x, t_{k+1}) - c_1''(x, t_k) - \tilde{\mathcal{L}}_4^* c_2(x, t_k), \quad x \in [1, 2), \\ &c_2(2, t_{k+1}) = 0, \\ &c_2(x, t_{k+1}) = b_2(x, t_{k+1}) \text{ on } [0, 1), \end{aligned}\right\} \tag{29}$$

and lastly, the functions b_3 and c_3 are the solutions of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_1 b_3(x, t_{k+1}) = -b_2''(x, t_{k+1}) - b_2''(x, t_k) - \tilde{\mathcal{L}}_1^* b_3(x, t_k), \quad x \in (0, 1), \\ &b_3(0, t_{k+1}) = 0, \quad b_3(1, t_{k+1}) = 0, \end{aligned}\right\} \tag{30}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_2 c_3(x,t_{k+1}) = -c_2''(x,t_{k+1}) - c_2''(x,t_k) - \tilde{\mathcal{L}}_2^* c_3(x,t_k), \quad x \in (1,2),\\ &c_3(1,t_{k+1}) = 0, \quad c_3(2,t_{k+1}) = 0,\\ &c_3(x,t_{k+1}) = b_3(x,t_{k+1}) \text{ on } [0,1). \end{aligned}\right\} \tag{31}$$

Lemma 3.6: *Let $\psi(x,t_{k+1})$ satisfies $\psi(1,t_{k+1}) \geq 0$ and $\tilde{\mathcal{L}}_3\psi(x,t_{k+1}) \leq 0$ for all $x \in [0,1)$. Then $\psi(x,t_{k+1}) \geq 0$ for all $x \in [0,1]$.*

Proof: Define $\psi(\alpha,t_{k+1}) = \min_{\alpha\in\bar{\Omega}_1} \psi(x,t_{k+1})$ for some $\alpha \in \bar{\Omega}_1$. Let $\psi(\alpha,t_{k+1}) < 0$, then $\alpha \neq 1$ and $\psi'(\alpha,t_{k+1}) = 0$. For $\alpha \in [0,1)$, we have $\tilde{\mathcal{L}}_3\psi(\alpha,t_{k+1}) = (\mu p\psi'(\alpha,t_{k+1}) - l\psi(\alpha,t_{k+1})) > 0$, which contradicts the assumption. Therefore, we can conclude that $\psi(\alpha,t_{k+1}) \geq 0$ which implies $\psi(x,t_{k+1}) \geq 0$ for all $x \in [0,1]$. ■

Lemma 3.7: *Let $b(x,t_{k+1})$ and $c(x,t_{k+1})$ be the solutions of* (24)–(31). *Then for $p_0\mu^2 \geq \gamma\,\varepsilon$ and $k = 0, 1, 2, 3$*

$$\|b^{(k)}\|_{\bar{\Omega}_1} \leq C\left(1 + \left(\frac{\varepsilon}{\mu}\right)^{3-k}\right), \quad \|c^{(k)}\|_{\bar{\Omega}_2} \leq C\left(1 + \left(\frac{\varepsilon}{\mu}\right)^{3-k}\right).$$

Proof: To establish the bounds for the derivatives of the smooth components b and c, we use the bounds for the derivatives of b_0, b_1, b_2, b_3, c_0, c_1, c_2 and c_3. Let's first establish the bounds for b_0 and c_0. To do this, we consider the decomposition of b_0 and c_0 as

$$\begin{aligned} b_0(x,t_{k+1};\mu) &= b_{0,0}(x,t_{k+1}) + \mu b_{0,1}(x,t_{k+1}) + \mu^2 b_{0,2}(x,t_{k+1})\\ &\quad + \mu^3 b_{0,3}(x,t_{k+1};\mu) \quad \text{for } x \in \bar{\Omega}_1,\\ c_0(x,t_{k+1};\mu) &= c_{0,0}(x,t_{k+1}) + \mu c_{0,1}(x,t_{k+1}) + \mu^2 c_{0,2}(x,t_{k+1})\\ &\quad + \mu^3 c_{0,3}(x,t_{k+1};\mu) \quad \text{for } x \in \bar{\Omega}_2, \end{aligned}$$

where the functions $b_{0,0}$ and $c_{0,0}$ are the solutions of

$$\left.\begin{aligned} &-(lb_{0,0})(x,t_{k+1}) = (mb_{0,0})(x,t_k) - r(x,t_{k+1})\psi_1(x-1,t_{k+1}) - r(x,t_k)\psi_1(x-1,t_k)\\ &\quad + g(x,t_k) + g(x,t_{k+1}), \quad x \in \bar{\Omega}_1,\\ &-(lc_{0,0})(x,t_{k+1}) + r(x,t_{k+1})b_{0,0}(x-1,t_{k+1}) = (mc_{0,0})(x,t_k) - r(x,t_k)b_{0,0}(x-1,t_k)\\ &\quad + g(x,t_k) + g(x,t_{k+1}), \quad x \in \bar{\Omega}_2, \end{aligned}\right\}$$

the functions $b_{0,1}$ and $c_{0,1}$ are the solutions of

$$\left.\begin{aligned} &(lb_{0,1})(x,t_{k+1}) = (pb_{0,0}')(x,t_{k+1}) + (pb_{0,0}' - mb_{0,1})(x,t_k), \quad x \in \bar{\Omega}_1,\\ &(lc_{0,1})(x,t_{k+1}) - r(x,t_{k+1})b_{0,1}(x-1,t_{k+1}) = (pc_{0,0}')(x,t_{k+1}) + (pc_{0,0}' - mc_{0,1})(x,t_k)\\ &\quad + r(x,t_k)b_{0,1}(x-1,t_k), \quad x \in \bar{\Omega}_2, \end{aligned}\right\}$$

the functions $b_{0,2}$ and $c_{0,2}$ are the solutions of

$$\left.\begin{aligned}&(lb_{0,2})(x,t_{k+1}) = (pb'_{0,1})(x,t_{k+1}) + (pb'_{0,1} - mb_{0,2})(x,t_k), \quad x \in \bar{\Omega}_1,\\ &(lc_{0,2})(x,t_{k+1}) - r(x,t_{k+1})b_{0,2}(x-1,t_{k+1}) = (pc'_{0,1})(x,t_{k+1}) + (pc'_{0,1} - mc_{0,2})(x,t_k)\\ &\quad + r(x,t_k)b_{0,2}(x-1,t_k), \quad x \in \bar{\Omega}_2,\end{aligned}\right\}$$

and lastly, the functions $b_{0,3}$ and $c_{0,3}$ are the solutions of

$$\left.\begin{aligned}&\tilde{\mathcal{L}}_3 b_{0,3}(x,t_{k+1}) = -(pb'_{0,2})(x,t_{k+1}) - (pb'_{0,2})(x,t_k) - \tilde{\mathcal{L}}_3^* b_{0,3}(x,t_k), \quad x \in [0,1),\\ &b_{0,3}(1,t_k) = 0,\end{aligned}\right\} \tag{32}$$

and

$$\left.\begin{aligned}&\tilde{\mathcal{L}}_4 c_{0,3}(x,t_{k+1}) = -(pc'_{0,2})(x,t_{k+1}) - (pc'_{0,2})(x,t_k) - \tilde{\mathcal{L}}_4^* c_{0,3}(x,t_k), \quad x \in [1,2),\\ &c_{0,3}(2,t_k) = 0,\\ &c_{0,3}(x,t_{k+1}) = b_{0,3}(x,t_{k+1}), \quad x \in [0,1).\end{aligned}\right\} \tag{33}$$

Since, the functions $l(x,t_{k+1})$, $m(x,t_{k+1})$, $r(x,t_{k+1})$, $\psi_1(x,t_{k+1})$, $\psi_2(x,t_{k+1})$, $\psi_3(x,t_{k+1})$, and $g(x,t_{k+1})$ are sufficiently smooth on their respective domain. Thus, we have

$$\begin{aligned}&\|b_{0,0}^{(k)}\|_{\bar{\Omega}_1} \le C, \quad \|c_{0,0}^{(k)}\|_{\bar{\Omega}_2} \le C, \quad \text{for } 0 \le k \le 10,\\ &\|b_{0,1}^{(k)}\|_{\bar{\Omega}_1} \le C, \quad \|c_{0,1}^{(k)}\|_{\bar{\Omega}_2} \le C, \quad \text{for } 0 \le k \le 9,\\ &\|b_{0,2}^{(k)}\|_{\bar{\Omega}_1} \le C, \quad \|c_{0,2}^{(k)}\|_{\bar{\Omega}_2} \le C, \quad \text{for } 0 \le k \le 8.\end{aligned} \tag{34}$$

Define $\psi^{\pm}(x,t_{k+1}) = C \pm b_{0,3}(x,t_{k+1})$, $x \in \bar{\Omega}_1$. Now, apply Lemma 3.6 to $\psi^{\pm}(x,t_{k+1})$ to obtain $\|b_{0,3}\|_{\bar{\Omega}1} \le C$. Then, we use (32) to get

$$\|b_{0,3}^{(k)}\|_{\bar{\Omega}_1} \le \frac{C}{\mu^k}, \quad \text{for } 0 \le k \le 8. \tag{35}$$

Similarly, consider the function $\psi^{\pm}(x,t_{k+1}) = C \pm c_{0,3}(x,t_{k+1})$, $x \in \bar{\Omega}_2$. Applying Lemma 3.6 to $\psi^{\pm}(x,t_{k+1})$ yields $\|c_{0,3}\|_{\bar{\Omega}_2} \le C$. Using (33), we have

$$\|c_{0,3}^{(k)}\|_{\bar{\Omega}_2} \le \frac{C}{\mu^k}, \quad \text{for } 0 \le k \le 8. \tag{36}$$

As defined in (24) and (25), choose $b_0(1,t_{k+1})$ and $c_0(2,t_{k+1})$ to be

$$\begin{aligned}b_0(1,t_{k+1}) &= (b_{0,0} + \mu b_{0,1} + \mu^2 b_{0,2})(1,t_{k+1}),\\ c_0(2,t_{k+1}) &= (c_{0,0} + \mu c_{0,1} + \mu^2 c_{0,2})(2,t_{k+1}).\end{aligned}$$

Therefore, using (34)–(36) to obtain

$$\|b_0^k\|_{\bar{\Omega}_1} \le C\left(1 + \frac{1}{\mu^{k-3}}\right), \quad \|c_0^k\|_{\bar{\Omega}_2} \le C\left(1 + \frac{1}{\mu^{k-3}}\right), \quad 0 \le k \le 8. \tag{37}$$

Now, decompose the functions b_1 and c_1 as

$$b_1(x,t_{k+1};\mu) = b_{1,0}(x,t_{k+1}) + \mu b_{1,1}(x,t_{k+1}) + \mu^2 b_{1,2}(x,t_{k+1};\mu), \quad x \in \bar{\Omega}_1,$$
$$c_1(x,t_{k+1};\mu) = c_{1,0}(x,t_{k+1}) + \mu c_{1,1}(x,t_{k+1}) + \mu^2 c_{1,2}(x,t_{k+1};\mu), \quad x \in \bar{\Omega}_2,$$

where the functions $b_{1,0}$ and $c_{1,0}$ are the solution of

$$\left.\begin{aligned}
&(lb_{1,0})(x,t_{k+1}) = b_0''(x,t_{k+1}) + b_0''(x,t_k) - (mb_{1,0})(x,t_k), \quad x \in \bar{\Omega}_1,\\
&(lc_{1,0})(x,t_{k+1}) - r(x,t_{k+1})c_{1,0}(x-1,t_{k+1}) = c_0''(x,t_{k+1}) + c_0''(x,t_k) - (mc_{1,0})(x,t_k)\\
&\quad + r(x,t_k)c_{1,0}(x-1,t_k), \quad x \in \bar{\Omega}_2,
\end{aligned}\right\}$$

the functions $b_{1,1}$ and $c_{1,1}$ are the solutions of

$$\left.\begin{aligned}
&(lb_{1,1})(x,t_{k+1}) = (pb_{1,0}')(x,t_{k+1}) + (pb_{1,0}' - mb_{1,1})(x,t_k), \quad x \in \bar{\Omega}_1,\\
&(lc_{1,1})(x,t_{k+1}) - r(x,t_{k+1})b_{1,1}(x-1,t_{k+1}) = (pc_{1,0}')(x,t_{k+1}) + (pc_{1,0}' - mc_{1,1})(x,t_k)\\
&\quad + r(x,t_k)b_{1,1}(x-1,t_k), \quad x \in \bar{\Omega}_2,
\end{aligned}\right\}$$

and the functions $b_{1,2}$ and $c_{1,2}$ are the solution of

$$\left.\begin{aligned}
&\tilde{\mathcal{L}}_3 b_{1,2}(x,t_{k+1}) = -(pb_{1,1}')(x,t_{k+1}) - (pb_{1,1}')(x,t_k) - \tilde{\mathcal{L}}_3^* b_{1,2}(x,t_k), \quad x \in [0,1),\\
&b_{1,2}(1,t_k) = 0,\\
&\tilde{\mathcal{L}}_4 c_{1,2}(x,t_{k+1}) = -(pc_{1,1}')(x,t_{k+1}) - (pc_{1,1}')(x,t_k) - \tilde{\mathcal{L}}_4^* c_{1,2}(x,t_k), \quad x \in [1,2),\\
&c_{1,2}(2,t_k) = 0,\\
&c_{1,2}(x,t_{k+1}) = b_{1,2}(x,t_{k+1}), \quad x \in [0,1).
\end{aligned}\right\}$$

To establish the bounds for the derivatives of b_1 and c_1, we use a similar argument as we did for b_0 and c_0. We apply Lemma 3.6 to suitable barrier functions and obtain the bounds for the derivatives of the components of b_1 and c_1 as follows

$$\|b_{1,0}^{(k)}\|_{\bar{\Omega}_1} \le C\left(1 + \frac{1}{\mu^{k-1}}\right), \quad \|c_{1,0}^{(k)}\|_{\bar{\Omega}_2} \le C\left(1 + \frac{1}{\mu^{k-1}}\right), \quad 0 \le k \le 6,$$
$$\|b_{1,1}^{(k)}\|_{\bar{\Omega}_1} \le C\left(1 + \frac{1}{\mu^{k}}\right), \quad \|c_{1,1}^{(k)}\|_{\bar{\Omega}_2} \le C\left(1 + \frac{1}{\mu^{k}}\right), \quad 0 \le k \le 5,$$
$$\|b_{1,2}^{(k)}\|_{\bar{\Omega}_1} \le \left(\frac{C}{\mu^{k+1}}\right), \quad \|c_{1,2}^{(k)}\|_{\bar{\Omega}_2} \le \left(\frac{C}{\mu^{k+1}}\right), \quad 0 \le k \le 5.$$

Therefore, we get the following bounds for the derivatives of b_1 and c_1

$$\|b_1^{(k)}\|_{\bar{\Omega}_1} \le C\left(1 + \frac{1}{\mu^{k-1}}\right), \quad \|c_1^{(k)}\|_{\bar{\Omega}_2} \le C\left(1 + \frac{1}{\mu^{k-1}}\right), \quad 0 \le k \le 5.$$

Similarly, we can establish the bounds for the derivatives of b_2 and c_2 as given below

$$\|b_2^{(k)}\|_{\bar{\Omega}_1} \le \left(\frac{C}{\mu^{k+1}}\right), \quad \|c_2^{(k)}\|_{\bar{\Omega}_2} \le \left(\frac{C}{\mu^{k+1}}\right), \quad 0 \le k \le 3. \tag{38}$$

From (30) and (31) and Lemma 2 of [12], we get the following bounds for b_3 and c_3

$$\|b_3\|_{\bar{\Omega}_1} \leq \left(\frac{C}{\mu^3}\right), \quad \|c_3\|_{\bar{\Omega}_2} \leq \left(\frac{C}{\mu^3}\right).$$

From Lemma 2.3, it follows that

$$\|b_3'\|_{\bar{\Omega}_1} \leq \left(\frac{C}{\varepsilon\mu^2}\right), \quad \|c_3'\|_{\bar{\Omega}_2} \leq \left(\frac{C}{\varepsilon\mu^2}\right), \tag{39}$$

and from (30) and (31), we have

$$\|b_3''\|_{\bar{\Omega}_1} \leq \left(\frac{C}{\varepsilon^2\mu}\right), \quad \|c_3''\|_{\bar{\Omega}_2} \leq \left(\frac{C}{\varepsilon^2\mu}\right). \tag{40}$$

Differentiating Equations (30) and (31) and using $p_0\mu^2 \geq \gamma\,\varepsilon$, we compute

$$\|b_3^{(3)}\|_{\bar{\Omega}_1} \leq \left(\frac{C}{\varepsilon^3}\right), \quad \|c_3^{(k)}\|_{\bar{\Omega}_2} \leq \left(\frac{C}{\varepsilon^3}\right). \tag{41}$$

Therefore, we can write

$$\|b_3^{(k)}\|_{\bar{\Omega}_1} \leq \left(\frac{C}{\varepsilon^k\mu^{3-k}}\right), \quad \|c_3^{(k)}\|_{\bar{\Omega}_2} \leq \left(\frac{C}{\varepsilon^k\mu^3 - k}\right), \quad k = 0, 1, 2, 3.$$

As defined in (16), choose $b(0, t_{k+1})$ and $b(1, t_{k+1})$ to be

$$\begin{aligned} b(0, t_{k+1}) &= b_0(0, t_{k+1}) + \varepsilon b_1(0, t_{k+1}) + \varepsilon^2 b_2(0, t_{k+1}), \\ b(1, t_{k+1}) &= b_0(1, t_{k+1}) + \varepsilon b_1(1, t_{k+1}), \end{aligned}$$

and from (17), choose $c(1, t_{k+1})$ and $c(2, t_{k+1})$ to be

$$\begin{aligned} c(1, t_{k+1}) &= c_0(1, t_{k+1}) + \varepsilon c_1(1, t_{k+1}) + \varepsilon^2 c_2(1, t_{k+1}), \\ c(2, t_{k+1}) &= c_0(2, t_{k+1}) + \varepsilon c_1(2, t_{k+1}). \end{aligned}$$

We now obtain the bounds for the functions b and c by using the estimates for their components b_0, b_1, b_2, b_3, c_0, c_1, c_2 and c_3 and get

$$\|b^{(k)}\|_{\bar{\Omega}_1} \leq C\left(1 + \left(\frac{\varepsilon}{\mu}\right)^{3-k}\right), \quad \|c^{(k)}\|_{\bar{\Omega}_2} \leq C\left(1 + \left(\frac{\varepsilon}{\mu}\right)^{3-k}\right). \qquad \blacksquare$$

3.3. Estimates for the singular components

Let us decompose the singular component $Z(x, t_{k+1})$ defined in (15) as

$$\begin{aligned} Z(x, t_{k+1}) &:= Z_L(x, t_{k+1}) + Z_R(x, t_{k+1}) \\ &:= (z_1^L + z_2^L)(x, t_{k+1}) + (z_1^R + z_2^R)(x, t_{k+1}), \end{aligned}$$

where the functions z_1^L and z_2^L satisfy

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_1 z_1^L(x, t_{k+1}) = 0, \quad x \in (0, 1), \\ &z_1^L(0, t_{k+1}) = U(0, t_{k+1}) - X(0, t_{k+1}) - j_1(\varepsilon, \mu), \quad z_1^L(1, t_{k+1}) = 0, \end{aligned}\right\} \tag{42}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_2 z_2^L(x, t_{k+1}) = 0, \quad x \in (1, 2), \\ &z_2^L(1, t_{k+1}) = k_1(\varepsilon, \mu) - j_2(\varepsilon, \mu), \quad z_2^L(2, t_{k+1}) = 0, \\ &z_2^L(x, t_{k+1}) = z_1^L(x, t_{k+1}), \quad x \in [0, 1), \end{aligned}\right\} \tag{43}$$

and the functions z_1^R and z_2^R satisfy

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_1 z_1^R(x, t_{k+1}) = 0, \quad x \in (0, 1), \\ &z_1^R(0, t_{k+1}) = j_1(\varepsilon, \mu), \quad z_1^R(1, t_{k+1}) = k_2(\varepsilon, \mu), \end{aligned}\right\} \tag{44}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_2 z_2^R(x, t_{k+1}) = 0, \quad x \in (1, 2), \\ &z_2^R(1, t_{k+1}) = j_2(\varepsilon, \mu), \quad z_2^R(2, t_{k+1}) = U(2, t_{k+1}) - X(2, t_{k+1}), \\ &z_2^R(x, t_{k+1}) = z_1^R(x, t_{k+1}), \quad x \in [0, 1). \end{aligned}\right\} \tag{45}$$

Here, $k_1(\varepsilon, \mu)$ and $k_2(\varepsilon, \mu)$ are constants to be chosen to satisfy the jump conditions at $x = 1$ given in (19) and (20). Moreover, constants $j_1(\varepsilon, \mu)$ and $j_2(\varepsilon, \mu)$ are to be chosen separately for the cases $p_0\mu^2 \le \gamma\,\varepsilon$ and $p_0\mu^2 \ge \gamma\,\varepsilon$ to satisfy the requirements for the bounds of the singular component.

Lemma 3.8: *Let $Z_L(x, t_{k+1})$ and $Z_R(x, t_{k+1})$ be the solutions of* (42)–(45). *Then for $p_0\mu^2 \ge \gamma\,\varepsilon$ and $k = 0, 1, 2, 3$*

$$\begin{aligned} &\|z_1^{L,(k)}\|_{\bar{\Omega}_1} \le C\left(\frac{\mu}{\varepsilon}\right)^k, \quad \|z_2^{L,(k)}\|_{\bar{\Omega}_2} \le C\left(\frac{\mu}{\varepsilon}\right)^k, \\ &\|z_1^{R,(k)}\|_{\bar{\Omega}_1} \le \frac{C}{\mu^k}, \quad \|z_2^{R,(k)}\|_{\bar{\Omega}_2} \le \frac{C}{\mu^k}. \end{aligned}$$

Proof: Since $U(0, t_{k+1})$ and $U(1, t_{k+1})$ are bounded by constants independent of ε and μ, $|j_1|$, $|j_2|$, $|k_1|$ and $|k_2|$ are also bounded by constants independent of ε and μ. Next, we use

Lemma 3.4 to compute

$$\|z_1^{L,(k)}(x,t_{k+1})\|_{\bar{\Omega}_1} \le C\left(\frac{\mu}{\varepsilon}\right)^k, \quad \|z_2^{L,(k)}(x,t_{k+1})\|_{\bar{\Omega}_2} \le C\left(\frac{\mu}{\varepsilon}\right)^k.$$

To find bounds for z_1^R and z_2^R, consider the following decomposition of z_1^R and z_2^R as

$$\begin{aligned} z_1^R(x,t_{k+1};\varepsilon,\mu) &= z_{1,0}^R(x,t_{k+1};\mu) + \varepsilon z_{1,1}^R(x,t_{k+1};\mu) + \varepsilon^2 z_{1,2}^R(x,t_{k+1};\mu) \\ &\quad + \varepsilon^3 z_{1,3}^R(x,t_{k+1};\varepsilon,\mu), \\ z_2^R(x,t_{k+1};\varepsilon,\mu) &= z_{2,0}^R(x,t_{k+1};\mu) + \varepsilon z_{2,1}^R(x,t_{k+1};\mu) + \varepsilon^2 z_{2,2}^R(x,t_{k+1};\mu) \\ &\quad + \varepsilon^3 z_{2,3}^R(x,t_{k+1};\varepsilon,\mu), \end{aligned} \tag{46}$$

where the functions $z_{1,0}^R$ and $z_{2,0}^R$ are the solutions of

$$\tilde{\mathcal{L}}_3 z_{1,0}^R(x,t_{k+1}) = 0, \quad x \in [0,1), \quad z_{1,0}^R(1,t_{k+1}) = k_2, \tag{47}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_4 z_{2,0}^R(x,t_{k+1}) = 0, \quad x \in [1,2), \quad z_{2,0}^R(2,t_{k+1}) = U(2,t_{k+1}) - X(2,t_{k+1}), \\ &z_{2,0}^R(x,t_{k+1}) = z_{1,0}^R(x,t_{k+1}), \quad x \in [0,1), \end{aligned}\right\} \tag{48}$$

the functions $z_{1,1}^R$ and $z_{2,1}^R$ are the solutions of

$$\tilde{\mathcal{L}}_3 z_{1,1}^R(x,t_{k+1}) = -z_{1,0}^{R,(2)}(x,t_{k+1}), \quad x \in [0,1), \quad z_{1,1}^R(1,t_{k+1}) = 0, \tag{49}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_4 z_{2,1}^R(x,t_{k+1}) = -z_{2,0}^{R,(2)}(x,t_{k+1}), \quad x \in [1,2), \quad z_{2,1}^R(2,t_{k+1}) = 0, \\ &z_{2,1}^R(x,t_{k+1}) = z_{1,1}^R(x,t_{k+1}), \quad x \in [0,1), \end{aligned}\right\} \tag{50}$$

the functions $z_{1,2}^R$ and $z_{2,2}^R$ are the solutions of

$$\tilde{\mathcal{L}}_3 z_{1,2}^R(x,t_{k+1}) = -z_{1,1}^{R,(2)}(x,t_{k+1}), \quad x \in [0,1), \quad z_{1,2}^R(1,t_{k+1}) = 0, \tag{51}$$

and

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_4 z_{2,2}^R(x,t_{k+1}) = -z_{2,1}^{R,(2)}(x,t_{k+1}), \quad x \in [1,2), \quad z_{2,2}^R(2,t_{k+1}) = 0, \\ &z_{2,2}^R(x,t_{k+1}) = z_{1,2}^R(x,t_{k+1}), \quad x \in [0,1), \end{aligned}\right\} \tag{52}$$

and lastly, the functions $z_{1,3}^R$ and $z_{2,3}^R$ are the solutions of

$$\left.\begin{aligned} &\tilde{\mathcal{L}}_1 z_{1,3}^R(x,t_{k+1}) = -z_{1,2}^{R,(2)}(x,t_{k+1}), \quad x \in (0,1),\ z_{1,3}^R(0,t_{k+1}) = 0,\ z_{1,3}^R(1,t_{k+1}) = 0, \\ &\tilde{\mathcal{L}}_2 z_{2,3}^R(x,t_{k+1}) = -z_{2,2}^{R,(2)}(x,t_{k+1}), \quad x \in (1,2),\ z_{2,3}^R(1,t_{k+1}) = 0,\ z_{2,3}^R(2,t_{k+1}) = 0, \\ &z_{2,3}^R(x,t_{k+1}) = z_{1,3}^R(x,t_{k+1}), \quad x \in [0,1). \end{aligned}\right\}$$

Define $\psi^{\pm}(x,t_{k+1}) = C \pm z_{1,0}^R(x,t_{k+1})$, $x \in [0,1]$ and apply Lemma 3.6 to the function $\psi^{\pm}(x,t_{k+1})$ to obtain $\|z_{1,0}^R(x,t_{k+1})\|_{\bar{\Omega}_1} \le C$. Now we use (47) to find

$$\|z_{1,0}^{R,(k)}\|_{\bar{\Omega}_1} \le \frac{C}{\mu^k}, \quad 0 \le k \le 5. \tag{53}$$

Similarly, it is easy to follow that

$$\|z_{2,0}^{R,(k)}\|_{\bar{\Omega}_2} \leq \frac{C}{\mu^k}, \quad 0 \leq k \leq 5. \tag{54}$$

Now, define $\psi^{\pm}(x, t_{k+1}) = C\|z_{1,0}^{R,(2)}\|_{\Omega_1} \pm z_{1,1}^{R}(x, t_{k+1})$, $x \in [0, 1]$ and apply Lemma 3.6 to the function $\psi^{\pm}$ we find $\|z_{1,1}^{R}\|_{\bar{\Omega}_1} \leq \frac{C}{\mu^2}$. From (49), we have

$$\|z_{1,1}^{R,(k)}\|_{\bar{\Omega}_1} \leq \frac{C}{\mu^{k+2}}, \quad 0 \leq k \leq 4. \tag{55}$$

Similarly, we can obtain the following bounds for $z_{1,2}^{R}$, $z_{2,1}^{R}$ and $z_{2,2}^{R}$

$$\|z_{1,2}^{R,(k)}\|_{\bar{\Omega}_1} \leq \frac{C}{\mu^{k+4}}, \quad 0 \leq k \leq 3, \tag{56}$$

$$\|z_{2,1}^{R,(k)}\|_{\bar{\Omega}_2} \leq \frac{C}{\mu^{k+2}}, \quad 0 \leq k \leq 4, \quad \|z_{2,2}^{R,(k)}\|_{\bar{\Omega}_2} \leq \frac{C}{\mu^{k+4}}, \quad 0 \leq k \leq 3. \tag{57}$$

Now, we apply Lemma 1 from [12] to the function $\psi^{\pm}(x, t_{k+1}) = C\|z_{1,2}^{R,(2)}\|_{\Omega_1} \pm z_{1,3}^{R}(x, t_{k+1})$, $x \in [0, 1]$, we get $\|z_{1,3}^{R}\|_{\bar{\Omega}_1} \leq \frac{C}{\mu^6}$. Using Lemma 3.4 to write

$$\|z_{1,3}^{R,(k)}\|_{\bar{\Omega}_1} \leq \frac{C}{(\sqrt{\varepsilon})^k \mu^6}\left(1 + \left(\frac{\mu}{\sqrt{\varepsilon}}\right)^k\right), \quad k = 1, 2, \tag{58}$$

$$\|z_{1,3}^{R,(3)}\|_{\bar{\Omega}_1} \leq \frac{C}{(\sqrt{\varepsilon})^3 \mu^6}\left(1 + \left(\frac{\mu}{\sqrt{\varepsilon}}\right)^3\right) + \frac{C}{\varepsilon \mu^7}. \tag{59}$$

Similarly, we can estimate the following bounds for $z_{2,3}^{R}$

$$\|z_{2,3}^{R,(k)}\|_{\bar{\Omega}_2} \leq \frac{C}{(\sqrt{\varepsilon})^k \mu^6}\left(1 + \left(\frac{\mu}{\sqrt{\varepsilon}}\right)^k\right), \quad k = 0, 1, 2, \tag{60}$$

$$\|z_{2,3}^{R,(3)}\|_{\bar{\Omega}_2} \leq \frac{C}{(\sqrt{\varepsilon})^3 \mu^6}\left(1 + \left(\frac{\mu}{\sqrt{\varepsilon}}\right)^3\right) + \frac{C}{\varepsilon \mu^7}, \quad k = 1, 2. \tag{61}$$

Since $p_0\mu^2 \geq \gamma\,\varepsilon$ and using (53)–(61) in (46), we get

$$\|z_1^{R,(k)}\|_{\bar{\Omega}_1} \leq \frac{C}{\mu^k} \quad \text{and} \quad \|z_2^{R,(k)}\|_{\bar{\Omega}_2} \leq \frac{C}{\mu^k}, \quad 0 \leq k \leq 3. \qquad \blacksquare$$

Lemma 3.9: *Let $Z_L(x, t_{k+1})$ and $Z_R(x, t_{k+1})$ be the solutions of* (42)–(45). *Then for $p_0\mu^2 \leq \gamma\,\varepsilon$ and $k = 0, 1, 2, 3$*

$$\|z_1^{L,(k)}(x, t_{k+1})\|_{\bar{\Omega}_1} \leq C\varepsilon^{\frac{-k}{2}}, \quad \|z_2^{L,(k)}(x, t_{k+1})\|_{\bar{\Omega}_2} \leq C\varepsilon^{\frac{-k}{2}},$$
$$\|z_1^{R,(k)}(x, t_{k+1})\|_{\bar{\Omega}_1} \leq C\varepsilon^{\frac{-k}{2}}, \quad \|z_2^{R,(k)}(x, t_{k+1})\|_{\bar{\Omega}_2} \leq C\varepsilon^{\frac{-k}{2}}.$$

Proof: Using Lemma 3.4, we note that z_1^L and z_2^L satisfy

$$\|z_1^{L,(k)}(x,t_{k+1})\|_{\bar{\Omega}_1} \leq C\varepsilon^{\frac{-k}{2}}, \quad \|z_2^{L,(k)}(x,t_{k+1})\|_{\bar{\Omega}_2} \leq C\varepsilon^{\frac{-k}{2}}, \quad 0 \leq k \leq 3.$$

For the bounds of z_1^R and z_2^R, choose $j_1(\varepsilon,\mu)$ and $j_2(\varepsilon,\mu)$ to be zero and using Lemma 3.4, we get

$$\|z_1^{R,(k)}\|_{\bar{\Omega}_1} \leq C\varepsilon^{\frac{-k}{2}} \quad \text{and} \quad \|z_2^{R,(k)}\|_{\bar{\Omega}_2} \leq C\varepsilon^{\frac{-k}{2}}, \quad 0 \leq k \leq 3. \qquad \blacksquare$$

The following Lemma establishes sharper estimates for the bounds of the singular components.

Lemma 3.10: *Let the layer components z_1^L, z_2^L, z_1^R and z_2^R be the solutions of* (42)–(45). *Then*

$$|z_1^L(x,t_{k+1})| \leq Ce^{-\theta_1 x},\ x \in [0,1], \quad |z_2^L(x,t_{k+1})| \leq Ce^{-\theta_1(x-1)},\ x \in [1,2],$$
$$|z_1^R(x,t_{k+1})| \leq Ce^{-\theta_2(1-x)},\ x \in [0,1], \quad |z_2^R(x,t_{k+1})| \leq Ce^{-\theta_2(2-x)},\ x \in [1,2],$$

where

$$\theta_1 = \begin{cases} \dfrac{\sqrt{\gamma p_0}}{\sqrt{\varepsilon}} & \text{if } \mu^2 \leq \dfrac{\gamma\varepsilon}{p_0}, \\ \dfrac{p_0\mu}{\varepsilon} & \text{if } \mu^2 \geq \dfrac{\gamma\varepsilon}{p_0}, \end{cases} \quad \text{and} \quad \theta_2 = \begin{cases} \dfrac{\sqrt{\gamma p_0}}{2\sqrt{\varepsilon}} & \text{if } \mu^2 \leq \dfrac{\gamma\varepsilon}{p_0}, \\ \dfrac{\gamma}{2\mu} & \text{if } \mu^2 \geq \dfrac{\gamma\varepsilon}{p_0}. \end{cases}$$

Proof: For proof, see [16]. $\blacksquare$

4. Spatial discretization

The solution of the problem exhibits boundary layers at $x = 0, x = 2$ and an interior layer at $x = 1$ [26]. To resolve the layers, we design the mesh to condense in the inner layer regions and to remain coarse in the outer regions, away from the layers. Consequently, to generate a piecewise-uniform meh $\bar{\mathbb{D}}_x^N$, we partition the interval $[0,2]$ into six subintervals as

$$[0,2] = [0,\beta_1] \cup [\beta_1, 1-\beta_2] \cup [1-\beta_2, 1] \cup [1, 1+\beta_1] \cup [1+\beta_1, 2-\beta_2] \cup [2-\beta_2, 2],$$

where β_1 and β_2 are the mesh transition parameters defined as

$$\beta_1 = \begin{cases} \min\left\{\dfrac{1}{4}, \dfrac{2\sqrt{\varepsilon}}{\sqrt{\gamma p_0}} \ln N\right\} & \text{if } \mu^2 \leq \dfrac{\gamma\varepsilon}{p_0}, \\ \min\left\{\dfrac{1}{4}, \dfrac{2\varepsilon}{p_0\mu} \ln N\right\} & \text{if } \mu^2 \geq \dfrac{\gamma\varepsilon}{p_0}, \end{cases} \quad \text{and}$$

$$\beta_2 = \begin{cases} \min\left\{\dfrac{1}{4}, \dfrac{2\sqrt{\varepsilon}}{\sqrt{\gamma p_0}} \ln N\right\} & \text{if } \mu^2 \leq \dfrac{\gamma\varepsilon}{p_0}, \\ \min\left\{\dfrac{1}{4}, \dfrac{2\mu}{\varepsilon} \ln N\right\} & \text{if } \mu^2 \geq \dfrac{\gamma\varepsilon}{p_0}. \end{cases}$$

We place $N/8$ mesh points each in intervals $[0, \beta_1]$, $[1-\beta_2, 1]$, $[1, 1+\beta_1]$, $[2-\beta_2, 2]$ and $N/4$ mesh points in intervals $[\beta_1, 1-\beta_2]$ and $[1+\beta_1, 2-\beta_2]$. Consequently, we obtain

$$\bar{\mathbb{D}}_x^N = \{x_i\}_0^N = \begin{cases} ih_1 & \text{for } i = 0, \ldots, N/8, \\ \beta_1 + \left(i - \dfrac{N}{8}\right) h_2 & \text{for } i = N/8+1, \ldots, 3N/8, \\ (1-\beta_2) + \left(i - \dfrac{3N}{8}\right) h_3 & \text{for } i = 3N/8+1, \ldots, N/2, \\ 1 + \left(i - \dfrac{N}{2}\right) h_1 & \text{for } i = N/2+1, \ldots, 5N/8, \\ (1+\beta_1) + \left(i - \dfrac{5N}{8}\right) h_2 & \text{for } i = 5N/8+1, \ldots, 7N/8, \\ (2-\beta_2) + \left(i - \dfrac{7N}{8}\right) h_3 & \text{for } i = 7N/8+1, \ldots, N, \end{cases}$$

where

$$h_i = \begin{cases} h_1 = \dfrac{4\beta_1}{N} & \text{for } i = 1, \ldots, N/8, N/2+1, \ldots, 5N/8, \\ h_2 = \dfrac{4}{N}(1-\beta_1-\beta_2) & \text{for } i = N/8+1, \ldots, 3N/8, 5N/8+1, \ldots, 7N/8, \\ h_3 = \dfrac{8\beta_2}{N} & \text{for } i = 3N/8+1, \ldots, N/2, 7N/8+1, \ldots, N. \end{cases}$$

The fully discrete problem on $\bar{\mathbb{D}}_x^N \times \mathbb{T}_t{}^M$ thus reads

$$\hat{\mathcal{L}}_n \tilde{U}(x_i, t_{k+1}) = \hat{\mathcal{G}}_n(x_i, t_{k+1}), \quad n = 1, 2, \tag{62}$$

where

$$\begin{aligned} \hat{\mathcal{L}}_1 \tilde{U}_{i,k+1} &= \varepsilon \delta_x^2 \tilde{U}_{i,k+1} + \mu p_{i,k+1} D_x^+ \tilde{U}_{i,k+1} - l_{i,k+1} \tilde{U}_{i,k+1} \quad \text{for } i = 1, \ldots, N/2-1, \\ \hat{\mathcal{L}}_2 \tilde{U}_{i,k+1} &= \varepsilon \delta_x^2 \tilde{U}_{i,k+1} + \mu p_{i,k+1} D_x^+ \tilde{U}_{i,k+1} - l_{i,k+1} \tilde{U}_{i,k+1} \\ &\quad + r_{i,k+1} \tilde{U}_{i-N/2,k+1} \quad \text{for } i = N/2, \ldots, N-1 \end{aligned}$$

and

$$\begin{aligned} \hat{\mathcal{G}}_1(x_i, t_{k+1}) &= -\varepsilon \delta_x^2 \tilde{U}_{i,k} - \mu p_{i,k} D_x^+ \tilde{U}_{i,k} + m_{i,k} \tilde{U}_{i,k} - r_{i,k} \psi_1(x_i - 1, t_k) \\ &\quad - r_{i,k+1} \psi_1(x_i - 1, t_{k+1}) + g_{i,k+1} + g_{i,k} \quad \text{for } i = 1, \ldots, N/2-1, \\ \hat{\mathcal{G}}_2(x_i, t_{k+1}) &= -\varepsilon \delta_x^2 \tilde{U}_{i,k} - \mu p_{i,k} D_x^+ \tilde{U}_{i,k} + m_{i,k} \tilde{U}_{i,k} - r_{i,k} \tilde{U}(x_i - 1, t_k) \\ &\quad + g_{i,k+1} + g_{i,k} \quad \text{for } i = N/2, \ldots, N-1 \end{aligned}$$

with

$$\begin{cases} \tilde{U}(x_i, t_{k+1}) = \psi_1(x_i - 1, t_{k+1}) & \text{for } i = 0, 1, \ldots N/2, \quad k = 0, \ldots, M-1, \\ \tilde{U}(x_i, t_0) = \psi_2(x_i, 0) & \text{for } i = 0, 1, \ldots, N, \\ \tilde{U}(x_N, t_{k+1}) = \psi_3(x_N, t_{k+1}) & \text{for } k = 0, \ldots, M-1. \end{cases}$$

The operator $\hat{\mathcal{L}}_n$ satisfies the following discrete minimum principle.

Lemma 4.1: *Let $Z_{i,k+1}$ be a mesh function such that $Z_{i,k+1} \geq 0$ for $i = \{0, N\}$, $\hat{\mathcal{L}}_1 Z_{i,k+1} \leq 0$ for $i = 1, \ldots, N/2 - 1$ and $\hat{\mathcal{L}}_2 Z_{i,k+1} \leq 0$ for $i = N/2, \ldots, N-1$. Then $Z_{i,k+1} \geq 0$ for $i = 0, 1, \ldots, N$.*

Proof: Choose $q^* \in \{0, 1, \ldots, N\}$ such that $Z_{q^*,k+1} = \min_{\bar{\mathbb{D}}_x^N \times \mathbb{T}_t^M} Z_{i,k+1}$. Suppose $Z_{q^*,k+1} < 0$ and it follows that $q^* \notin \{0, N\}$, $Z_{q^*+1,k+1} - Z_{q^*,k+1} \geq 0$ and $Z_{q^*,k+1} - Z_{q^*-1,k+1} \leq 0$.

Case I: For $q^* \in \{1, 2, \ldots, N/2 - 1\}$

$$\begin{aligned}
\hat{\mathcal{L}}_1 Z_{q^*,k+1} &= \varepsilon \delta_x^2 Z_{q^*,k+1} + \mu p_{q^*,k+1} D_x^+ Z_{q^*,k+1} - l_{q^*,k+1} Z_{q^*,k+1} \\
&= \frac{2\varepsilon}{h_{q^*} + h_{q^*+1}} \left\{ \frac{Z_{q^*+1,k+1} - Z_{q^*,k+1}}{h_{q^*+1}} - \frac{Z_{q^*,k+1} - Z_{q^*-1,k+1}}{h_{q^*}} \right\} \\
&\quad + \mu p_{q^*,k+1} \left\{ \frac{Z_{q^*+1,k+1} - Z_{q^*,k+1}}{h_{q^*+1}} \right\} - l_{q^*,k+1} Z_{q^*,k+1} \\
&> 0.
\end{aligned}$$

Case II: For $q^* \in \{N/2, \ldots, N-1\}$

$$\begin{aligned}
\hat{\mathcal{L}}_2 Z_{q^*,k+1} &= \varepsilon \delta_x^2 Z_{q^*,k+1} + \mu p_{q^*,k+1} D_x^+ Z_{q^*,k+1} - l_{q^*,k+1} Z_{q^*,k+1} + r_{q^*,k+1} Z_{q^*-N/2,k+1} \\
&\geq (r_{q^*,k+1} - l_{q^*,k+1}) Z_{q^*,k+1} \\
&> 0.
\end{aligned}$$

The required result follows from a contradiction. ■

Consequently, we can prove the following stability estimate for discrete operator $\hat{\mathcal{L}}_n$.

Lemma 4.2: *Let $Z_{i,k+1}$ be the solution of* (62). *Then*

$$|Z_{i,k+1}| \leq \max \left\{ |Z_{0,k+1}|, |Z_{N,k+1}|, \|\hat{\mathcal{L}}_1 Z_{i,k+1}\|, \|\hat{\mathcal{L}}_2 Z_{i,k+1}\| \right\}.$$

Proof: For proof, see [28]. ■

5. Error estimates

Let us decompose $\tilde{U}_{i,k+1}$ into smooth and singular components to obtain parameter uniform error estimates. We write $\tilde{U}_{i,k+1} := \tilde{X}_{i,k+1} + \tilde{Z}_{i,k+1}$, where

$$\tilde{X}(x_i, t_{k+1}) = \begin{cases} \tilde{X}_1(x_i, t_{k+1}) & \text{for } 0 \leq i \leq N/2 - 1, \\ \tilde{X}_2(x_i, t_{k+1}) & \text{for } N/2 \leq i \leq N, \end{cases}$$

and

$$\tilde{Z}(x_i, t_{k+1}) = \begin{cases} \tilde{Z}_1(x_i, t_{k+1}) & \text{for } 0 \leq i \leq N/2 - 1, \\ \tilde{Z}_2(x_i, t_{k+1}) & \text{for } N/2 \leq i \leq N. \end{cases}$$

The smooth component $\tilde{X}$ satisfies

$$\left.\begin{aligned}
&\hat{\mathcal{L}}_1\tilde{X}_1(x_i,t_{k+1}) = \hat{\mathcal{G}}_1(x_i,t_{k+1}) \quad \text{for } i\in\{1,2,\ldots,N/2-1\},\\
&\tilde{X}_1(x_0,t_{k+1}) = b(0,t_{k+1}), \quad \tilde{X}_1(x_{N/2},t_{k+1}) = b(1,t_{k+1}),\\
&\hat{\mathcal{L}}_2\tilde{X}_2(x_i,t_{k+1}) = \hat{\mathcal{G}}_2(x_i,t_{k+1}) \quad \text{for } i\in\{N/2+1,\ldots,N-1\},\\
&\tilde{X}_2(x_{N/2},t_{k+1}) = c(1,t_{k+1}), \quad \tilde{X}_2(x_N,t_{k+1}) = c(2,t_{k+1}),\\
&\tilde{X}_2(x_i-1,t_{k+1}) = \tilde{X}_1(x_{i-N/2},t_{k+1}), \quad \text{for } i\in\{N/2,\ldots,N-1\},
\end{aligned}\right\} \tag{63}$$

and the singular component $\tilde{Z}$ satisfies

$$\left.\begin{aligned}
&\hat{\mathcal{L}}_1\tilde{Z}_1(x_i,t_{k+1}) = 0 \quad \text{for } i\in\{1,2,\ldots,N/2-1\},\\
&\tilde{Z}_1(x_0,t_{k+1}) = z_1(0,t_{k+1}), \quad \tilde{Z}_1(x_{N/2},t_{k+1}) = Z_1(1,t_{k+1}),\\
&\hat{\mathcal{L}}_2\tilde{Z}_2(x_i,t_{k+1}) = 0 \quad \text{for } i\in\{N/2+1,\ldots,N-1\},\\
&\tilde{Z}_2(x_{N/2},t_{k+1}) = z_2(1,t_{k+1}), \quad \tilde{Z}_2(x_N,t_{k+1}) = z_2(2,t_{k+1}).
\end{aligned}\right\} \tag{64}$$

Moreover, the error $e_{i,k+1}$ is given by

$$e_{i,k+1} = \tilde{U}_{i,k+1} - U_{i,k+1} = (\tilde{X}_{i,k+1} - X_{i,k+1}) + (\tilde{Z}_{i,k+1} - Z_{i,k+1}).$$

Lemma 5.1: *Let* $X(x_i,t_{k+1})$ *and* $\tilde{X}_{i,k+1}$ *be the solutions of* (14), (16), (17) *and* (63), *respectively. Then*

$$|\tilde{X}_{i,k+1} - X(x_i,t_{k+1})| \le CN^{-1}, \quad 0\le i\le N.$$

Proof: Consider

$$\begin{aligned}
|\tilde{X}(x_i,t_{k+1}) - X(x_i,t_{k+1})| &= |(\tilde{X}_1+\tilde{X}_2)(x_i,t_{k+1}) - (b+c)(x_i,t_{k+1})|\\
&= |(\tilde{X}_1 - b)(x_i,t_{k+1}) + (\tilde{X}_2 - c)(x_i,t_{k+1})|\\
&\le |(\tilde{X}_1 - b)(x_i,t_{k+1})| + |(\tilde{X}_2 - c)(x_i,t_{k+1})|. \qquad (65)
\end{aligned}$$

For $i = 1,\ldots,N/2-1$

$$\begin{aligned}
&|\hat{\mathcal{L}}_1(\tilde{X}_1 - b)(x_i,t_{k+1})|\\
&\quad = |\hat{\mathcal{L}}_1\tilde{X}_1(x_i,t_{k+1}) - \hat{\mathcal{L}}_1 b(x_i,t_{k+1})|\\
&\quad \le \varepsilon\left|\left(\delta_x^2 - \frac{\mathrm{d}^2}{\mathrm{d}x^2}\right)b(x_i,t_{k+1})\right| + \mu|p(x_i,t_{k+1})|\left|\left(D_x^+ - \frac{\mathrm{d}}{\mathrm{d}x}\right)b(x_i,t_{k+1})\right|\\
&\quad \le C\max_{1\le i\le \frac{N}{2}-1} h_i(\varepsilon\|b\|_3 + \mu\|b\|_2).
\end{aligned}$$

Case I: For $p_0\mu^2 \le \gamma\,\varepsilon$. Lemma 3.5 leads to

$$|\hat{\mathcal{L}}_1(\tilde{X}_1 - b)(x_i,t_{k+1})| \le CN^{-1}.$$

Case II: For $p_0\mu^2 \geq \gamma\,\varepsilon$. Lemma 3.7 leads to

$$|\hat{\mathcal{L}}_1(\tilde{X}_1 - b)(x_i, t_{k+1})| \leq CN^{-1}.$$

Define $\psi^{\pm}(x_i, t_{k+1}) = CN^{-1} \pm (\tilde{X}_1 - b)(x_i, t_{k+1})$, $0 \leq i \leq N/2 - 1$ for sufficiently large C. Then $\psi^{\pm}(x_0, t_{k+1}) \geq 0$, $\psi^{\pm}(x_{N/2}, t_{k+1}) \geq 0$ and $\hat{\mathcal{L}}_1\psi^{\pm}(x_i, t_{k+1}) \leq 0$. Now, using Lemma 4.1, we obtain

$$|(\tilde{X}_1 - b)(x_i, t_{k+1})| \leq CN^{-1}, \quad 0 \leq i \leq N/2.$$

A similar argument for $(\tilde{X}_2 - c)$ leads us to the following estimates

$$|(\tilde{X}_2 - c)(x_i, t_{k+1})| \leq CN^{-1}, \quad N/2 \leq i \leq N.$$

Therefore, from (65), it follows that $|\tilde{X}(x_i, t_{k+1}) - X(x_i, t_{k+1})| \leq CN^{-1}$, $0 \leq i \leq N$. ■

Let us decompose $\tilde{Z}_1(x, t_{k+1})$ and $\tilde{Z}_2(x, t_{k+1})$ further as $\tilde{Z}_1(x_i, t_{k+1}) := (\tilde{Z}_1^L + \tilde{Z}_1^R)(x_i, t_{k+1})$ and $\tilde{Z}_2(x_i, t_{k+1}) := (\tilde{Z}_2^L + \tilde{Z}_2^R)(x_i, t_{k+1})$, where $\tilde{Z}_1^L$ and $\tilde{Z}_2^L$ satisfy

$$\left.\begin{aligned}
&\hat{\mathcal{L}}_1\tilde{Z}_1^L(x_i, t_{k+1}) = 0, \quad 0 < i < N/2,\\
&\tilde{Z}_1^L(x_0, t_{k+1}) = z_1^L(0, t_{k+1}), \quad \tilde{Z}_1^L(x_{N/2}, t_{k+1}) = 0,\\
&\hat{\mathcal{L}}_2\tilde{Z}_2^L(x_i, t_{k+1}) = 0, \quad N/2 < i < N,\\
&\tilde{Z}_2^L(x_{N/2}, t_{k+1}) = z_2^L(1, t_{k+1}), \quad \tilde{Z}_2^L(x_N, t_{k+1}) = 0,\\
&\tilde{Z}_2^L(x_{N/2} - 1, t_{k+1}) = \tilde{Z}_1^L(x_{i-N/2}, t_{k+1}), \quad N/2 \leq i < N,
\end{aligned}\right\} \tag{66}$$

and $\tilde{Z}_1^R$ and $\tilde{Z}_2^R$ satisfy

$$\left.\begin{aligned}
&\hat{\mathcal{L}}_1\tilde{Z}_1^R(x_i, t_{k+1}) = 0, \quad 0 < i < N/2,\\
&\tilde{Z}_1^R(x_0, t_{k+1}) = z_1^R(0, t_{k+1}), \quad \tilde{Z}_1^R(x_{N/2}, t_{k+1}) = z_1^R(1, t_{k+1}),\\
&\hat{\mathcal{L}}_2\tilde{Z}_2^R(x_i, t_{k+1}) = 0, \quad N/2 < i < N,\\
&\tilde{Z}_2^R(x_{N/2}, t_{k+1}) = z_2^R(1, t_{k+1}), \quad \tilde{Z}_2^R(x_N, t_{k+1}) = z_2^R(2, t_{k+1}),\\
&\tilde{Z}_2^R(x_i - 1, t_{k+1}) = \tilde{Z}_1^R(x_{i-N/2}, t_{k+1}), \quad 0 \leq i < N/2.
\end{aligned}\right\} \tag{67}$$

Define the mesh functions

$$I_{1,i} = \begin{cases} \displaystyle\prod_{j=1}^{i}(1 + \theta_1 h_j)^{-1}, & 1 < i \leq N/2,\\ 1, & i = 0, \end{cases} \quad \text{and}$$

$$I_{2,i} = \begin{cases} \displaystyle\prod_{j=i+1}^{N/2}(1 + \theta_2 h_j)^{-1}, & 0 \leq i < N/2,\\ 1, & i = N/2, \end{cases}$$

where θ_1 and θ_2 are defined in Lemma 3.10. For $i = N/8, \ldots, N/2$

$$\begin{aligned} I_{1,i} &\le I_{1,N/8} \\ &= \prod_{j=1}^{N/8} (1 + \theta_1 h_{N/8})^{-1} \\ &= (1 + \theta_1 h_{N/8})^{-N/8} \\ &= \left(1 + \theta_1 \frac{8\beta_1}{N}\right)^{-N/8} \\ &\le (1 + 8N^{-1} \ln N)^{-N/8} \le CN^{-1}, \end{aligned} \tag{68}$$

and for $i = 0, \ldots, 3N/8$

$$\begin{aligned} I_{2,i} &\le I_{2,3N/8} \\ &= \left(1 + \theta_2 \frac{8\beta_2}{N}\right)^{-N/8} \\ &\le (1 + 8N^{-1} \ln N)^{-N/8} \\ &\le CN^{-1}. \end{aligned} \tag{69}$$

Lemma 5.2: *Let $\tilde{Z}_1^L$, $\tilde{Z}_2^L$, $\tilde{Z}_1^R$ and $\tilde{Z}_2^R$ be the solutions of* (66) *and* (67). *Then*

$$|\tilde{Z}_1^L(x_i, t_{k+1})| \le CI_{1,i}, \quad |\tilde{Z}_1^R(x_i, t_{k+1})| \le CI_{2,i}, \quad 0 \le i \le N/2 \quad \text{and}$$
$$|\tilde{Z}_2^L(x_i, t_{k+1})| \le CI_{1,i-\frac{N}{2}}, \quad |\tilde{Z}_2^R(x_i, t_{k+1})| \le CI_{2,i-\frac{N}{2}}, \quad N/2 \le i \le N.$$

Proof: For $i = 0, \ldots N/2$, define $\mathcal{Q}^{\pm}(x_i, t_{k+1}) = |\tilde{Z}_1^L(x_0, t_{k+1})| I_{1,i} \pm \tilde{Z}_1^L(x_i, t_{k+1})$. Then, $\mathcal{Q}^{\pm}(x_0, t_{k+1}) = |\tilde{Z}_1^L(x_0, t_{k+1})| \pm \tilde{Z}_1^L(x_0, t_{k+1}) \ge 0$ and $\mathcal{Q}^{\pm}(x_{N/2}, t_{k+1}) \ge 0$. Also, note that $\hat{\mathcal{L}}_1 I_{1,i} = (\varepsilon \delta_s^2 + \mu p D_s^+ - l) I_{1,i} \le 0$ which implies $\hat{\mathcal{L}}_1 \mathcal{Q}^{\pm}(x_i, t_{k+1}) \le 0$. Using Lemma 4.1, we have

$$|\tilde{Z}_1^L(x_i, t_{k+1})| \le CI_{1,i}, \quad 0 \le i \le N/2.$$

A similar argument for $\tilde{Z}_1^R$, $\tilde{Z}_2^L$ and $\tilde{Z}_2^R$ leads to the following estimates

$$\begin{aligned} &|\tilde{Z}_1^R(x_i, t_{k+1})| \le CI_{2,i}, \quad 0 \le i \le N/2, \\ &|\tilde{Z}_2^L(x_i, t_{k+1})| \le CI_{1,i-N/2}, \quad N/2 \le i \le N, \\ &|\tilde{Z}_2^R(x_i, t_{k+1})| \le CI_{2,i-N/2}, \quad N/2 \le i \le N. \end{aligned}$$

■

To estimate the error bound for $(\tilde{Z} - Z) = (\tilde{Z}_L - Z_L) + (\tilde{Z}_R - Z_R)$, we compute the error bound for $(\tilde{Z}_L - Z_L)$ and $(\tilde{Z}_R - Z_R)$, separately. For this, define $\tilde{Z}_L$ and $\tilde{Z}_R$ as

$$\tilde{Z}_L(x_i, t_{k+1}) = \begin{cases} \tilde{Z}_1^L, & 0 \le i \le N/2 - 1, \\ \tilde{Z}_2^L, & N/2 \le i \le N, \end{cases} \qquad \text{and}$$

$$\tilde{Z}_R(x_i, t_{k+1}) = \begin{cases} \tilde{Z}_1^R, & 0 \le i \le N/2 - 1, \\ \tilde{Z}_2^R, & N/2 \le i \le N. \end{cases}$$

Therefore

$$(\tilde{Z}_L - Z_L)(x_i, t_{k+1}) = \begin{cases} (\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1}), & 0 \le i \le N/2 - 1, \\ (\tilde{Z}_2^L - z_2^L)(x_i, t_{k+1}), & N/2 \le i \le N, \end{cases}$$

and

$$(\tilde{Z}_R - Z_R)(x_i, t_{k+1}) = \begin{cases} (\tilde{Z}_1^R - z_1^R)(x_i, t_{k+1}), & 0 \le i \le N/2 - 1, \\ (\tilde{Z}_2^R - z_2^R)(x_i, t_{k+1}), & N/2 \le i \le N. \end{cases}$$

Lemma 5.3: *Let Z_L, Z_R be the solutions of* (42)–(45) *and $\tilde{Z}_L, \tilde{Z}_R$ be the solutions of* (66)–(67), *respectively. Then*

$$|(\tilde{Z}_L - Z_L)(x_i, t_{k+1})| \le \begin{cases} CN^{-1} \ln N & \text{if } p_0 \mu^2 \le \gamma\, \varepsilon, \\ CN^{-1} (\ln N)^2 & \text{if } p_0 \mu^2 \ge \gamma\, \varepsilon, \end{cases}$$

$$|(\tilde{Z}_R - Z_R)(x_i, t_{k+1})| \le \begin{cases} CN^{-1} \ln N & \text{if } p_0 \mu^2 \le \gamma\, \varepsilon, \\ CN^{-1} \ln N & \text{if } p_0 \mu^2 \ge \gamma\, \varepsilon. \end{cases}$$

Proof: Using (68), (69) and Lemma 5.2, we have

$$|\tilde{Z}_1^L(x_i, t_{k+1})| \le CN^{-1}, \quad \text{for } N/8 \le i \le N/2, \tag{70}$$

$$|\tilde{Z}_1^R(x_i, t_{k+1})| \le CN^{-1}, \quad \text{for } 0 \le i \le 3N/8, \tag{71}$$

$$|\tilde{Z}_2^L(x_i, t_{k+1})| \le CN^{-1}, \quad \text{for } 5N/8 \le i \le N, \tag{72}$$

$$|\tilde{Z}_2^R(x_i, t_{k+1})| \le CN^{-1}, \quad \text{for } N/2 \le i \le 7N/8. \tag{73}$$

Case I: For $p_0 \mu^2 \le \gamma\, \varepsilon$ and $1 \le i \le N/8 - 1$

$$\begin{aligned} |\hat{\mathcal{L}}_1(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| &= |\hat{\mathcal{L}}_1 \tilde{Z}_1^L(x_i, t_{k+1}) - \hat{\mathcal{L}} z_1^L(x_i, t_{k+1})| \\ &= |\hat{\mathcal{L}}_1 \tilde{Z}_1^L(x_i, t_{k+1}) - \tilde{\mathcal{L}}_1 z_1^L(x_i, t_{k+1})| \\ &\le \varepsilon \left(\delta_x^2 - \frac{d^2}{dx^2} \right) |z_1^L(x_i, t_{k+1})| + \mu \left(D_x^+ - \frac{d}{dx} \right) |z_1^L(x_i, t_{k+1})| \\ &\le Ch_i \left(\varepsilon |z_1^L|_3 + \mu |z_1^L|_2 \right) \\ &\le Ch_i \left(\frac{1}{\sqrt{\varepsilon}} + \frac{\mu}{\sqrt{\varepsilon}} \right) \\ &\le \frac{Ch_i}{\sqrt{\varepsilon}}. \end{aligned}$$

In case $\beta_1 = \dfrac{1}{4}$, we have $\dfrac{1}{4} \le \dfrac{2\sqrt{\varepsilon}}{\sqrt{\gamma\, p_0}} \ln N$, which implies $\dfrac{1}{\sqrt{\varepsilon}} \le C \ln N$. Thus

$$|\hat{\mathcal{L}}_1(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| \le CN^{-1} \ln N.$$

In case $\beta_1 = \dfrac{2\sqrt{\varepsilon}}{\sqrt{\gamma\, p_0}} \ln N$, we compute

$$|\hat{\mathcal{L}}_1(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| \leq \frac{Ch_i}{\sqrt{\varepsilon}} \leq \frac{C}{\sqrt{\varepsilon}} \frac{8\beta_1}{N} \leq CN^{-1} \ln N.$$

Let us next consider $\phi^{\pm}(x_i, t_{k+1}) = CN^{-1} \ln N \pm (\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})$ for $0 \leq i \leq N/8$. Then $\phi^{\pm}(x_i, t_{k+1}) \geq 0$ for $i = \{0, N/8\}$ and $\hat{\mathcal{L}}_1 \phi^{\pm}(x_i, t_{k+1}) \leq 0$. Using Lemma 4.1, we get

$$|(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| \leq CN^{-1} \ln N, \quad 0 \leq i \leq N/8. \tag{74}$$

For $N/8 \leq i \leq N/2 - 1$, using (70) and Lemma 3.10, we calculate that

$$|(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1}) \leq |\tilde{Z}_1^L(x_i, t_{k+1})| + |z_1^L(x_i, t_{k+1})| \leq CN^{-1}. \tag{75}$$

Thus from (74) and (75), we get

$$|(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1}) \leq CN^{-1} \ln N, \quad 0 \leq i \leq N/2 - 1.$$

Similarly, we can compute that

$$|(\tilde{Z}_2^L - z_2^L)(x_i, t_{k+1})| \leq CN^{-1} \ln N, \quad N/2 \leq i \leq N.$$

Case II: For $p_0 \mu^2 \geq \gamma\, \varepsilon$ and $1 \leq i \leq N/8 - 1$

$$\begin{aligned} |\hat{\mathcal{L}}_1(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| &= |\hat{\mathcal{L}}_1 \tilde{Z}_1^L(x_i, t_{k+1}) - \tilde{\mathcal{L}}_1 z_1^L(x_i, t_{k+1})| \\ &\leq Ch_i \left(\varepsilon |z_1^L|_3 + \mu |z_1^L|_2\right) \\ &\leq Ch_i \left(\frac{\mu^3}{\varepsilon^2} + \frac{\mu^3}{\varepsilon^2}\right) \\ &\leq Ch_i \mu \left(\frac{\mu}{\varepsilon}\right)^2. \end{aligned}$$

In case $\beta_1 = \dfrac{1}{4}$, we have $\dfrac{1}{4} \leq \dfrac{2\varepsilon}{p_0 \mu} \ln N$, which implies $\dfrac{\mu}{\varepsilon} \leq C \ln N$. Thus

$$|\hat{\mathcal{L}}_1(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| \leq CN^{-1} (\ln N)^2.$$

If $\beta_1 = \dfrac{2\varepsilon}{p_0 \mu} \ln N$, then we compute

$$|\hat{\mathcal{L}}_1(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| \leq Ch_i \mu \left(\frac{\mu}{\varepsilon}\right)^2 \leq C \frac{\mu^2}{\varepsilon} N^{-1} \ln N.$$

Consider,

$$\begin{aligned} \phi^{\pm}(x_i, t_{k+1}) = CN^{-1} + CN^{-1} \ln N (\beta_1 - x_i) \frac{\mu}{\varepsilon} \\ \pm (\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1}), \quad 0 \leq i \leq N/8. \end{aligned}$$

Then, $\hat{\mathcal{L}}_1\phi^{\pm}(x_i, t_{k+1}) \leq 0$ for $1 \leq i \leq N/8 - 1$ and $\phi^{\pm}(x_i, t_{k+1}) \geq 0$ for $i = \{0, N/8\}$. Using Lemma 4.1 to get

$$|(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1})| \leq CN^{-1} + CN^{-1} \ln N \beta_1 \frac{\mu}{\varepsilon}$$
$$\leq CN^{-1}(\ln N)^2, \quad 0 \leq i \leq N/8. \tag{76}$$

For $N/8 \leq i \leq N/2 - 1$, it follows from (70) and Lemma 3.10 that

$$|(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1}) \leq CN^{-1}. \tag{77}$$

Thus, from (76) and (77)

$$|(\tilde{Z}_1^L - z_1^L)(x_i, t_{k+1}) \leq CN^{-1}(\ln N)^2, \quad 0 \leq i \leq N/2 - 1. \tag{78}$$

Similarly, for $N/2 \leq i \leq N$, we obtain

$$|(\tilde{Z}_2^L - z_2^L)(x_i, t_{k+1})| \leq CN^{-1}(\ln N)^2.$$

Thus,

$$|(\tilde{Z}_L - Z_L)(x_i, t_{k+1})| \leq \begin{cases} CN^{-1} \ln N & \text{if } p_0\mu^2 \leq \gamma\,\varepsilon, \\ CN^{-1}(\ln N)^2 & \text{if } p_0\mu^2 \geq \gamma\,\varepsilon. \end{cases}$$

Following the similar steps and argument for $(\tilde{Z}_R - Z_R)$, it is straightforward to establish that for $p_0\mu^2 \leq \gamma\,\varepsilon$ and $p_0\mu^2 \geq \gamma\,\varepsilon$,

$$|(\tilde{Z}_R - Z_R)(x_i, t_{k+1})| \leq CN^{-1} \ln N, \quad 0 \leq i \leq N. \qquad \blacksquare$$

Theorem 5.4: *Let* $\tilde{U}(x_i, t_{k+1})$ *and* $U(x_i, t_{k+1})$ *be the solutions of* (62) *and* (10)*, respectively. Then for* $0 \leq i \leq N$

$$\left|(\tilde{U} - U)(x_i, t_{k+1})\right| \leq C \begin{cases} N^{-1} \ln N & \text{if } p_0\mu^2 \leq \gamma\,\varepsilon \\ N^{-1}(\ln N)^2 & \text{if } p_0\mu^2 \geq \gamma\,\varepsilon. \end{cases}$$

Proof: The proof follows from Lemmas 5.1 and 5.3. $\blacksquare$

Finally, we combine (13) and Theorem 5.4 to obtain the principle convergence estimate that reads.

Theorem 5.5: *Let* u *and* $\tilde{U}$ *be the solutions of the continuous problem* (1) *and the discrete problem* (62)*, respectively. Then*

$$\left|\tilde{U}(x_i, t_{k+1}) - u(x_i, t_{k+1})\right| \leq C \begin{cases} (\Delta t^2 + N^{-1} \ln N) & \text{if } p_0\mu^2 \leq \gamma\,\varepsilon \\ (\Delta t^2 + N^{-1}(\ln N)^2) & \text{if } p_0\mu^2 \geq \gamma\,\varepsilon \end{cases}$$

for $0 \leq i \leq N$ *and* $0 \leq k \leq M$.

Table 1. Maximum absolute error and order of convergence for Example 6.1 for different values of μ, M and N when $\varepsilon = 2^{-10}$.

μ	$M = N = 32$	64	128	256	512	1024
2^{-11}	2.8356e−03	1.3289e−03	5.8495e−04	2.2566e−04	7.5573e−05	2.4680e−05
	1.0934	1.1838	1.3742	1.5782	1.6145	1.6390
2^{-13}	2.8186e−03	1.3309e−03	5.8587e−04	2.2603e−04	7.5589e−05	2.4635e−05
	1.0826	1.1837	1.3741	1.5803	1.6175	1.6422
2^{-15}	2.8143e−03	1.3314e−03	5.8611e−04	2.2612e−04	7.5593e−05	2.4624e−05
	1.0798	1.1837	1.3741	1.5808	1.6182	1.6487
2^{-17}	2.8133e−03	1.3315e−03	5.8617e−04	2.2615e−04	7.5594e−05	2.4621e−05
	1.0792	1.1837	1.3740	1.5809	1.6184	1.6525
2^{-19}	2.8130e−03	1.3315e−03	5.8618e−04	2.2615e−04	7.5595e−05	2.4620e−05
	1.0791	1.1836	1.3741	1.5809	1.6185	1.6588
2^{-21}	2.8129e−03	1.3315e−03	5.8618e−04	2.2616e−04	7.5595e−05	2.4620e−05
	1.0789	1.1836	1.3740	1.5810	1.6185	1.6588
2^{-23}	2.8129e−03	1.3315e−03	5.8619e−04	2.2616e−04	7.5595e−05	2.4620e−05
	1.0789	1.1836	1.3740	1.5810	1.6185	1.6588
2^{-25}	2.8129e−03	1.3315e−03	5.8619e−04	2.2616e−04	7.5595e−05	2.4620e−05
	1.0789	1.1836	1.3740	1.5810	1.6185	1.6588

Table 2. Maximum absolute error and order of convergence for Example 6.1 for different values of ε, M and N when $\mu = 2^{-6}$.

ε	$M = N = 32$	64	128	256	512	1024
2^{-14}	5.8061e−03	4.6143e−03	3.2823e−03	2.0873e−03	1.2587e−03	7.3043e−04
	0.3315	0.4914	0.6530	0.7297	0.7851	0.8412
2^{-16}	6.9481e−03	5.2511e−03	3.5537e−03	2.3566e−03	1.4543e−03	8.5388e−04
	0.4040	0.5632	0.5926	0.6963	0.7682	0.8403
2^{-18}	7.1592e−03	5.7505e−03	4.0863e−03	2.6675e−03	1.6513e−03	9.5415e−04
	0.3161	0.4928	0.6153	0.7094	0.7913	0.8577
2^{-20}	7.4927e−03	5.9608e−03	4.2997e−03	2.7905e−03	1.7028e−03	9.0641e−04
	0.3299	0.4712	0.6236	0.7126	0.7730	0.8524
2^{-22}	7.5787e−03	6.0219e−03	4.3625e−03	2.8267e−03	1.7237e−03	1.0087e−03
	0.3317	0.4650	0.6260	0.7136	0.7730	0.8571
2^{-24}	7.5956e−03	6.0352e−03	4.3788e−03	2.8362e−03	1.7291e−03	1.0120e−03
	0.3317	0.4628	0.6265	0.7139	0.7728	0.8592
2^{-26}	7.6080e−03	6.0390e−03	4.3822e−03	2.8385e−03	1.7305e−03	1.0129e−03
	0.3332	0.4626	0.6265	0.7139	0.7726	0.8611
2^{-28}	7.6156e−03	6.0421e−03	4.3830e−03	2.8389e−03	1.7308e−03	1.0131e−03
	0.3339	0.4631	0.6266	0.7138	0.7726	0.8583

6. Numerical illustrations

In this section, we consider two model problems, present numerical results using the proposed method, and verify the theoretical estimates numerically.

Example 6.1: Consider the following two-parameter singularly perturbed problem:

$$\begin{cases} [\varepsilon u_{xx} + \mu(1+x)u_x - (4+\sin x)u - u_t](x,t) + u(x-1,t) \\ \quad = -e^x, \quad (x,t) \in (0,2) \times (0,2], \\ u(x,t) = 5 + t^2, \quad (x,t) \in \Gamma_1, \\ u(x,t) = 5, \quad (x,t) \in \Gamma_2, \\ u(x,t) = 5, \quad (x,t) \in \Gamma_3. \end{cases}$$

Table 3. Maximum absolute error and order of convergence for Example 6.2 for different values of μ, M and N when $\varepsilon = 2^{-20}$.

μ	$M = N = 32$	64	128	256	512	1024
2^{-18}	1.024e−02	5.3320e−03	1.9856e−03	6.7329e−04	2.1566e−04	6.4616e−05
	1.0478	1.4250	1.5602	1.6424	1.7388	1.8226
2^{-20}	1.0914e−02	5.2812e−03	1.9729e−03	6.6492e−04	2.1370e−04	6.4556e−05
	1.0471	1.4205	1.5690	1.6375	1.7270	1.8192
2^{-22}	1.0887e−02	5.2688e−03	1.9697e−03	6.6289e−04	2.1332e−04	6.4510e−05
	1.0470	1.4195	1.5711	1.6357	1.7254	1.8093
2^{-24}	1.0880e−02	5.2657e−03	1.9689e−03	6.6238e−04	2.1323e−04	6.4506e−05
	1.0470	1.4192	1.5716	1.6352	1.7249	1.8046
2^{-26}	1.0878e−02	5.2649e−03	1.9687e−03	6.6225e−04	2.1320e−04	6.4489e−05
	1.0469	1.4191	1.5717	1.6351	1.7251	1.7963
2^{-28}	1.0878e−02	5.2647e−03	1.9686e−03	6.6222e−04	2.1320e−04	6.4489e−05
	1.0469	1.4191	1.5718	1.6351	1.7251	1.7963
2^{-30}	1.0878e−02	5.2647e−03	1.9686e−03	6.6221e−04	2.1320e−04	6.4489e−05
	1.0469	1.4191	1.5718	1.6350	1.7251	1.7963
2^{-32}	1.0878e−02	5.2646e−03	1.9686e−03	6.6221e−04	2.1320e−04	6.4489e−05
	1.0469	1.4191	1.5718	1.6350	1.7251	1.7963

Table 4. Maximum absolute error and order of convergence for Example 6.2 for different values of ε, M and N when $\mu = 2^{-8}$.

ε	$M = N = 32$	64	128	256	512	1024
2^{-16}	1.2594e−02	7.7325e−03	3.7780e−03	1.5195e−03	6.0382e−04	2.2666e−04
	0.7037	1.0333	1.3140	1.3314	1.4135	1.5628
2^{-18}	1.3692e−02	8.3363e−03	3.8621e−03	1.5195e−03	6.0025e−04	2.2632e−04
	0.7158	1.1096	1.3462	1.3321	1.4072	1.5646
2^{-20}	1.4047e−02	8.5373e−03	3.8621e−03	1.5182e−03	5.0182e−04	2.2602e−04
	0.7184	1.1444	1.3470	1.3350	1.4129	1.5673
2^{-22}	1.4143e−02	8.5913e−03	3.8622e−03	1.5467e−03	6.0154e−04	2.2583e−04
	0.7191	1.1535	1.3467	1.3351	1.4134	1.5771
2^{-24}	1.4167e−02	8.6051e−03	3.8633e−03	1.5172e−03	6.0131e−04	2.2566e−04
	0.7127	1.1554	1.3484	1.3352	1.4140	1.5842
2^{-26}	1.4174e−02	8.6086e−03	3.8531e−03	1.5119e−03	5.9672e−04	2.2322e−04
	0.7193	1.1598	1.3497	1.3412	1.4186	1.5886
2^{-28}	1.4175e−02	8.6096e−03	3.8526e−03	1.5112e−03	5.9626e−04	2.2209e−04
	0.7193	1.1601	1.3501	1.3417	1.4248	1.5904
2^{-30}	1.4175e−02	8.6096e−03	3.8526e−03	1.5112e−03	5.9626e−04	2.2209e−04
	0.7193	1.1601	1.3501	1.3417	1.4248	1.5904

Example 6.2: Consider the following two-parameter singularly perturbed problem:

$$\begin{cases} [\varepsilon u_{xx} + \mu(2 + x + t)u_x - (2 + xt)u - u_t](x,t) + u(x-1,t) = (e^{t^2} - 1)(1 + xt), \\ (x,t) \in (0,2) \times (0,2], \\ u(x,t) = 0, \quad (x,t) \in \Gamma_1, \\ u(x,t) = 0, \quad (x,t) \in \Gamma_2, \\ u(x,t) = 0, \quad (x,t) \in \Gamma_3. \end{cases}$$

The exact solutions of the above examples are not known for comparison. Therefore, the double mesh principle [20] is used to estimate the maximum absolute error and rate of convergence. The maximum absolute error ($E_\varepsilon^{N,\Delta t}$) and order of convergence ($R_\varepsilon^{N,\Delta t}$) are

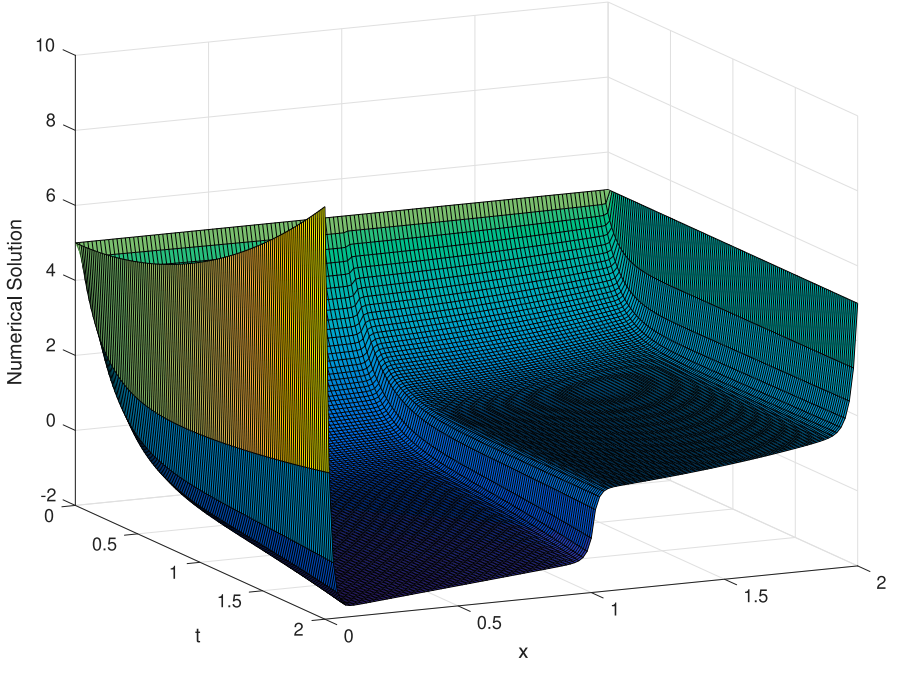


Figure 1. Numerical solution of Example 6.1 with $M = N = 128$, $\mu = 2^{-22}$ and $\varepsilon = 2^{-17}$.

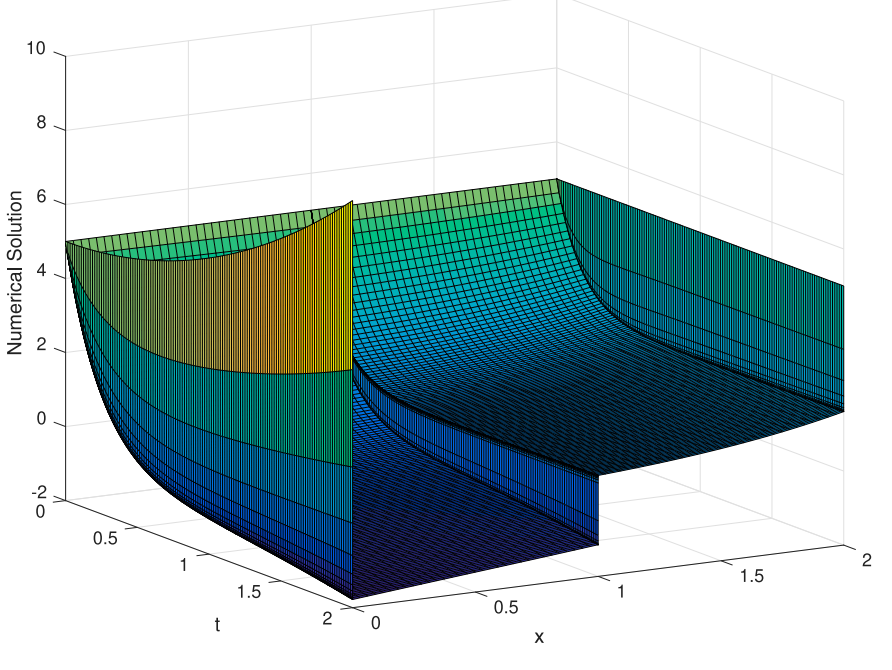


Figure 2. Numerical solution of Example 6.1 with $M = N = 128$, $\mu = 2^{-6}$ and $\varepsilon = 2^{-14}$.

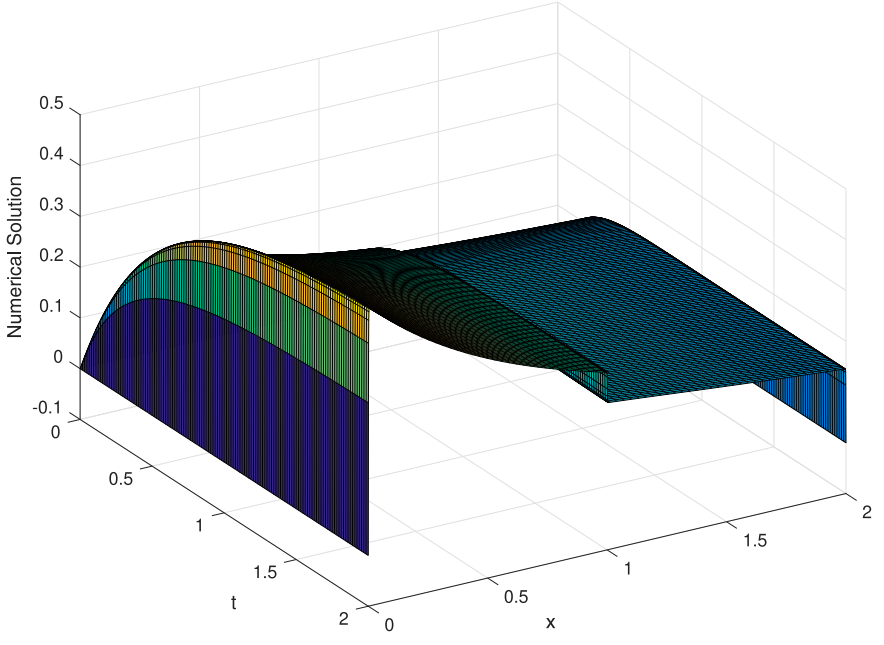


Figure 3. Numerical solution of Example 6.2 with $M = N = 128$, $\mu = 2^{-18}$ and $\varepsilon = 2^{-20}$.

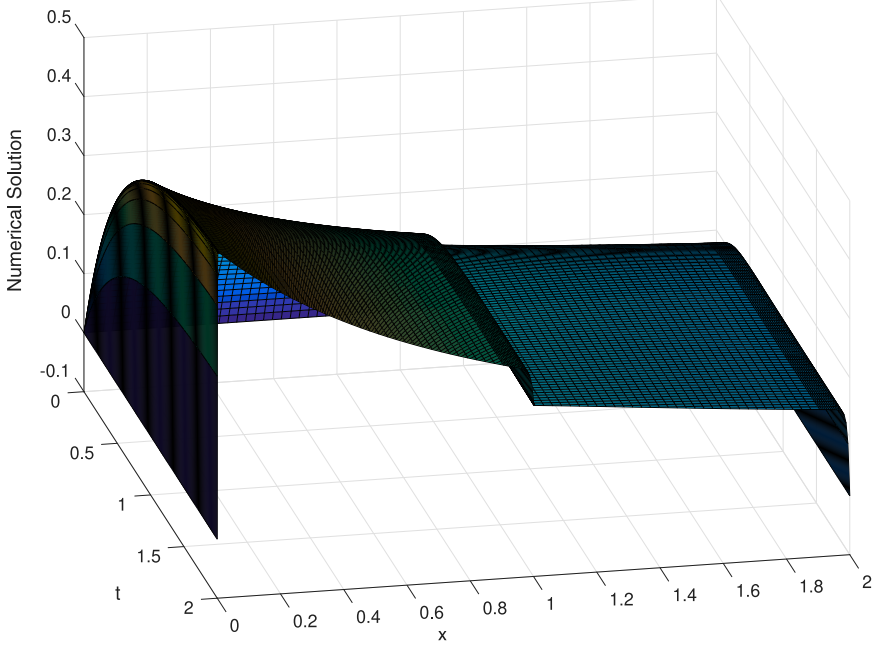


Figure 4. Numerical solution of Example 6.2 with $M = N = 128$, $\mu = 2^{-8}$ and $\varepsilon = 2^{-18}$.

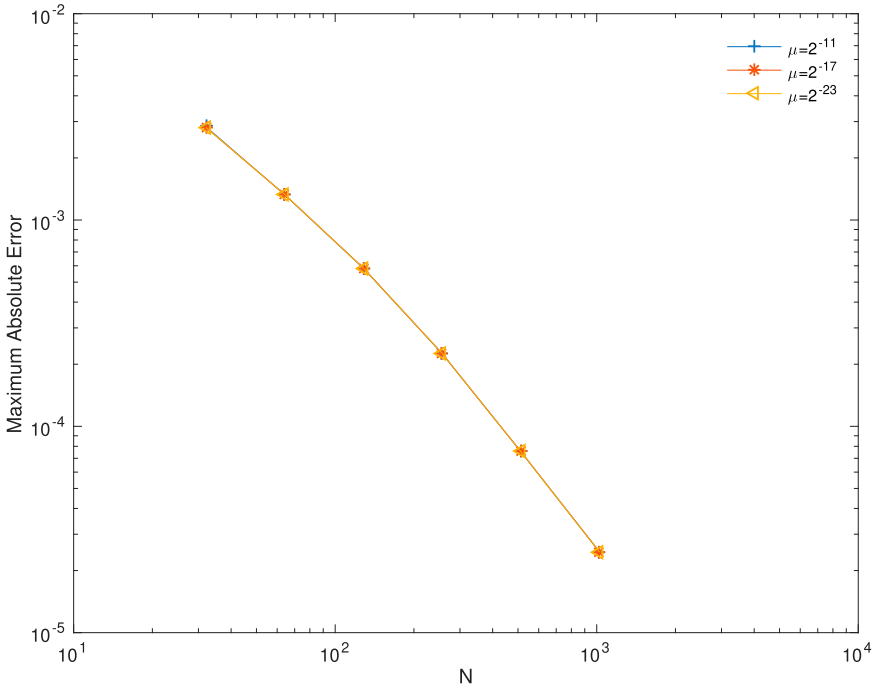


Figure 5. Error plot for Example 6.1 for Case 1 when $\varepsilon = 2^{-10}$.

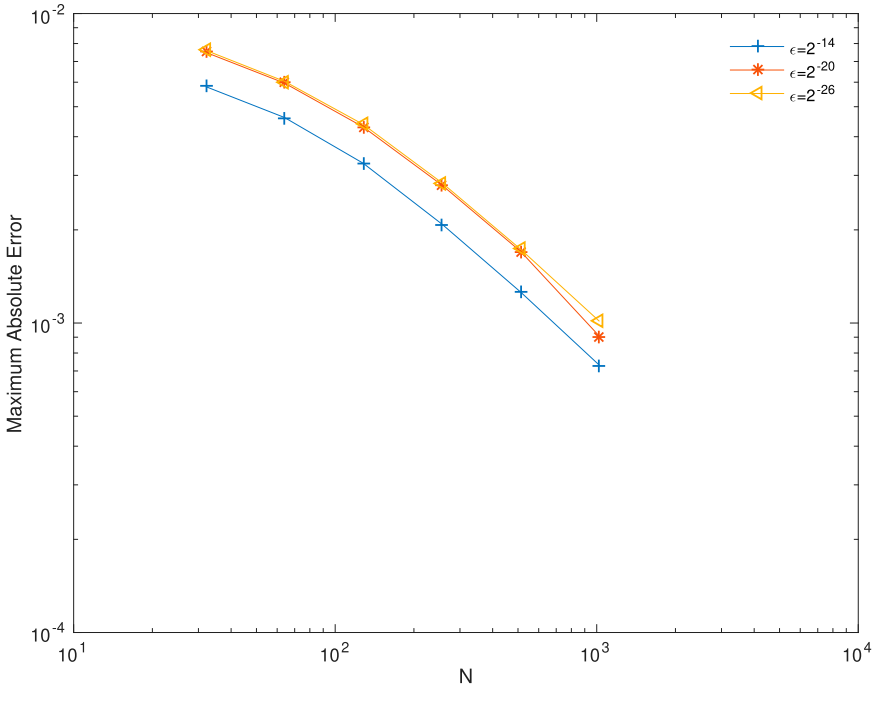


Figure 6. Error plot for Example 6.1 for Case 2 when $\mu = 2^{-6}$.

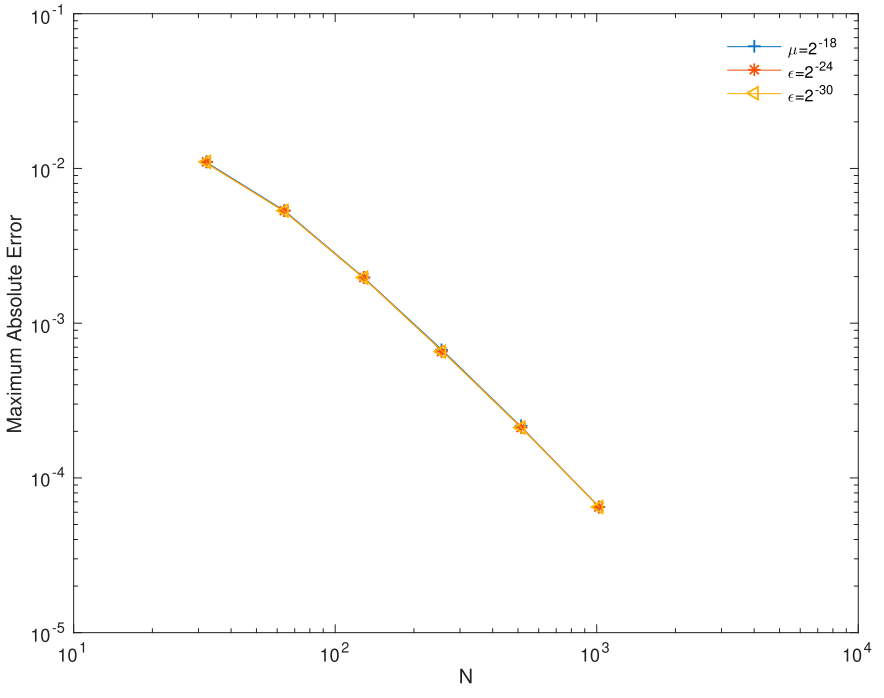


Figure 7. Error plot for Example 6.2 for Case 1 when $\varepsilon = 2^{-20}$.

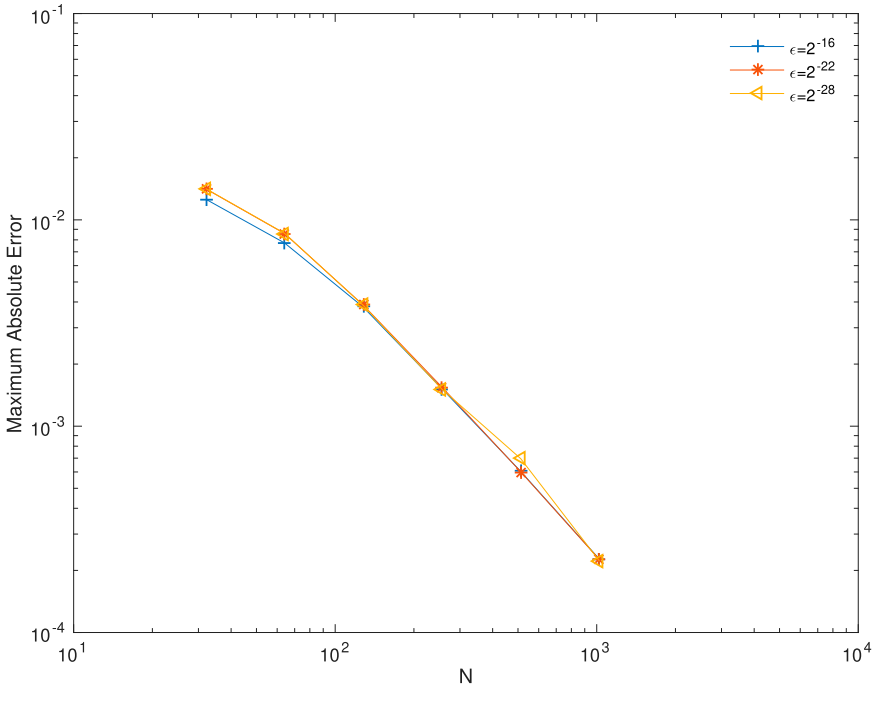


Figure 8. Error plot for Example 6.2 for Case 2 when $\mu = 2^{-8}$.

defined as

$$E_{\varepsilon}^{N,\Delta t} := \max \left| U_{N,\Delta t}(x_i, t_{k+1}) - \tilde{U}_{2N,\Delta t/2}(x_i, t_{k+1}) \right| \quad \text{and}$$

$$R_{\varepsilon}^{N,\Delta t} := \log_2 \left(\frac{E_{\varepsilon}^{N,\Delta t}}{E_{\varepsilon}^{2N,\Delta t/2}} \right).$$

Here, $U_{N,\Delta t}(x_i, t_{k+1})$ and $\tilde{U}_{2N,\Delta t/2}(x_i, t_{k+1})$ denotes the numerical solutions on $\bar{\mathbb{D}}_x^N \times \mathbb{T}_t{}^M$ and $\bar{\mathbb{D}}_x^{2N} \times \mathbb{T}_t{}^{2M}$, respectively.

The maximum absolute error ($E_{\varepsilon}^{N,\Delta t}$) and corresponding order of convergence ($R_{\varepsilon}^{N,\Delta t}$) for Examples 6.1 and 6.2 are tabulated for different values of ε, μ, M, and N in Tables 1–4. Moreover, the log-log plot of the maximum absolute error can be had from Figures 5–8. It is evident that the errors decrease monotonically as N increases.

The presence of both interior and boundary layers is apparent from the surface plots of the numerical solution of Examples 6.1 and 6.2 are displayed in Figures 1–4, for different values of ε and μ.

7. Concluding remark

This work proposed a higher-order difference method on a non-uniform mesh to solve singularly perturbed two-parameter parabolic problems with a large shift. The problem is characterized by the presence of two small parameters, leading to multiple timescales within the system. The interplay between fast and slow dynamics and non-local effects induced by shift gives rise to the boundary and interior layers. These layers are regions of rapid variation in the solution, and they can pose significant challenges for numerical solutions. The method combines an upwind difference scheme over a non-uniform mesh in space and a Crank-Nicolson scheme in time. A rigorous theoretical analysis is performed to establish the proposed method's consistency, stability, and convergence. The method is unconditionally stable, free from directional bias and parameters uniformly convergent with second-order accuracy in time and first-order accuracy in the space variable. Numerical results and illustrations are presented for two test problems which support theoretical estimates.

Author Contribution

All the authors contributed equally to all sections of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This document is the result of the research project MTR/2021/000117 funded by Science and Engineering Research Board, Department of Science & Technology, Government of India.

Data availability

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

[1] K. Aarthika, V. Shanthi, and H. Ramos, A computational approach for a two-parameter singularly perturbed system of partial differential equations with discontinuous coefficients, Appl. Math. Comput. 434 (2022), p. 127409.

[2] D. Avijit and S. Natesan, Convergence analysis of a fully-discrete fem for singularly perturbed two-parameter parabolic pde, Math. Comput. Simul. 197 (2022), pp. 185–206.

[3] M.A. Ayele, A.A. Tiruneh, and G.A. Derese, Fitted cubic spline scheme for two-parameter singularly perturbed time-delay parabolic problems, Results Appl. Math. 18 (2023), p. 100361.

[4] M. Brdar and H. Zarin, A singularly perturbed problem with two parameters on a Bakhvalov type mesh, J. Comput. Appl. Math. 292 (2016), pp. 307–319.

[5] M. Brdar, H. Zarin, and L. Teofanov, A singularly perturbed problem with two parameters in two dimensions on graded meshes, Comput. Math. Appl. 72 (2016), pp. 2582–2603.

[6] M. Chandru, P. Das, and H. Ramos, Numerical treatment of two parameter singularly perturbed parabolic convection diffusion problems with nonsmooth data, Math. Methods Appl. Sci. 41 (2018), pp. 5359–5387.

[7] M. Chandru, T. Prabha, and V. Shanthi, A parameter robust higher order numerical method for singularly perturbed two parameter problems with non-smooth data, J. Comput. Appl. Math. 309 (2017), pp. 11–27.

[8] J. Chen and R.E. O'Malley, On the asymptotic solution of a two-parameter boundary value problem ofchemical reactor theory, SIAM J. Appl. Math. 26 (1974), pp. 717–729.

[9] C. Clavero, J.L. Gracia, and J.C. Jorge, High-order numerical methods for one-dimensional parabolic singularly perturbed problems with regular layers, Numer. Methods Partial Differ. Equations 21 (2004), pp. 148–169.

[10] R.C. DiPrima, Asumptotic methods for an infinitely long slider squeeze-film bearing, Trans. ASME Ser. F. J. Lub. Tech. 90 (1968), pp. 173–183.

[11] L. Govindarao, J. Mohapatra, and S.R. Sahu, Uniformly convergent numerical method for singularly perturbed two parameter time delay parabolic problem, Int. J. Appl. Comput. Math. 5 (2019), p. 91.

[12] J.L. Gracia, E. O'Riordan, and M.L. Pickett, A parameter robust second order numerical method for a singularly perturbed two-parameter problem, Appl. Numer. Math. 56 (2006), pp. 962–980.

[13] V. Gupta, M.K. Kadalbajoo, and R.K. Dubey, A parameter-uniform higher order finite difference scheme for singularly perturbed time-dependent parabolic problem with two small parameters, Int. J. Comput. Math. 96(3) (2019), pp. 474–499.

[14] V. Gupta, M. Kumar, and S. Kumar, Higher order numerical approximation for time dependent singularly perturbed differential-difference convection-diffusion equations, Numer. Methods. Partial. Differ. Equ. 34(1) (2018), pp. 357–380.

[15] M. Kadalbajoo and A.S. Yadaw, Parameter-uniform finite element method for two-parameter singularly perturbed parabolic reaction-diffusion problems, Int. J. Comput. Methods 9 (2012), p. 1250047.

[16] S.S. Kalaiselvan, J.J.H. Miller, and V. Sigamani, A parameter uniform numerical method for a singularly perturbed two-parameter delay differential equation, Appl. Numer. Math. 145 (2019), pp. 90–110.

[17] A. Kaushik, Nonstandard perturbation approximation and travelling wave solutions of nonlinear reaction diffusion equations, Numer. Methods Partial Differ. Equations 24 (2008), pp. 217–238.

[18] A. Kaushik, Singular perturbation analysis of bistable differential equation arising in the nerve pulse propagation, Nonlinear Anal. 9 (2008), pp. 2106–2127.

[19] O.A. Ladyzhenskaya, V.A. Solonnikov, and N.N. Ural'ceva, Linear and Quasilinear Equations of Parabolic Type. Translated From the Russian by S. Smith, Translations of Mathematical Monographs, Vol. 23, American Mathematical Society, Providence, 1968.
[20] T. Linß, Layer Adapted Meshes for Reaction Convection Diffusion Problems, Springer, Berlin, 2010.
[21] N.T. Negero, Fitted cubic spline in tension difference scheme for two-parameter singularly perturbed delay parabolic partial differential equations, Partial Differ. Equ. Appl. Math. 8 (2023), p. 100530.
[22] N.T. Negero, A parameter-uniform efficient numerical scheme for singularly perturbed time-delay parabolic problems with two small parameters, Partial Differ. Equ. Appl. Math. 7 (2023), p. 100518.
[23] R.E. O'Malley, Two-parameter singular perturbation problems for second order equations, J. Math. Mech. 16 (1967), pp. 1143–1164.
[24] R.E. O'Malley, Introduction to Singular Perturbations, Academic Press, New York, 1974.
[25] R.E. O'Malley, Singular Perturbation Methods for Ordinary Differential Equations, Springer, New York, 1991.
[26] E. O'Riordan, M.L. Pickett, and G.I. Shishkin, Parameter-uniform finite difference schemes for singularly perturbed parabolic diffusion-convection-reaction problems, Math. Comp. 75 (2006), pp. 1135–1154.
[27] S.K. Sahoo and V. Gupta, Parameter robust higher-order finite difference method for convection-diffusion problem with time delay, Numer. Methods. Partial. Differ. Equ. 39(6) (2023), pp. 4145–4173.
[28] N. Sharma and A. Kaushik, A uniformly convergent difference method for singularly perturbed parabolic partial differential equations with large delay and integral boundary condition, J. Appl. Math. Comput. 69 (2023), pp. 1071–1093.
[29] S.C. Shiralashetti and S.I. Hanaji, Hermite wavelet based numerical method for the solution of two parameters singularly perturbed non-linear benjamina-bona-mohany equation, Sci. African 12 (2021), p. e00770.
[30] R. Shiromani, V. Shanthi, and H. Ramos, A computational method for a two-parameter singularly perturbed elliptic problem with boundary and interior layers, Math. Comput. Simul. 206 (2023), pp. 40–64.
[31] Sumit, S. Kumar, Kuldeep, and M. Kumar, A robust numerical method for a two-parameter singularly perturbed time delay parabolic problem, Comput. Appl. Math. 39 (2020), p. 209.
[32] J. Zhang and L. Yanhui, High-order finite element method on a bakhvalov-type mesh for a singularly perturbed convection–diffusion problem with two parameters, Appl. Math. Comput. 397 (2021), p. 125953.

Review Article

IT Medical Team
https://www.itmedicalteam.pl/

Health Science Journal
ISSN 1791-809X

2024
Vol. 18 No. 1: 1096

Addressing Health Disparities through Community-Based Public Health Initiatives

Naima Tiwari*

Department of Public Health, Delhi Technological University, India

***Corresponding author:**
Naima Tiwari

nai_tiwari23@gmail.com

Department of Public Health, Delhi Technological University, India

Citation: Tiwari N (2024) Addressing Health Disparities through Community-Based Public Health Initiatives. Health Sci J. Vol. 18 No. 1: 1096.

Abstract

This research article explores the critical role of community-based public health initiatives in addressing health disparities. Health disparities, defined as differences in health outcomes between various populations, remain a significant challenge worldwide, affecting individuals across socioeconomic, racial, ethnic, and geographic lines. Despite advancements in medical technology and healthcare delivery, disparities persist, highlighting the need for innovative approaches to promote health equity. Community-based public health initiatives leverage local resources, knowledge, and partnerships to develop targeted interventions that address the unique needs of diverse populations. This article examines the principles, strategies, and outcomes of community-based public health initiatives in reducing health disparities and promoting population health. Drawing upon examples from various communities, the article highlights successful approaches, challenges, and opportunities for collaboration among stakeholders. Additionally, it discusses the importance of cultural competence, community engagement, and sustainability in designing and implementing effective interventions. By fostering partnerships, empowering communities, and advocating for policy change, community-based public health initiatives offer promising solutions to mitigate health disparities and advance health equity.

Keywords: Public health; Health disparities; Community-based initiatives; Health equity; Community engagement; Population health

Received: 02-Jan-2024, Manuscript No. iphsj-24-14541; **Editor assigned:** 05-Jan-2024, Preqc No. iphsj-24-14541; **Reviewed:** 19-Jan-2024, QC No. iphsj-24-14541; **Revised:** 26-Jan-2024, Manuscript No. iphsj-24-14541; **Published:** 31-Jan-2024, DOI: 10.36648/1791-809X.18.1.1096

Introduction

Health disparities, defined as differences in health outcomes between various populations, persist as a significant public health challenge globally. Despite advancements in medical science and healthcare delivery, disparities in access to care, health outcomes, and health behaviours continue to affect individuals across socioeconomic, racial, ethnic, and geographic lines [1]. These disparities not only undermine the well-being of affected populations but also pose challenges to achieving health equity a fundamental principle of public health. Addressing health disparities requires multifaceted strategies that go beyond clinical interventions to address the social, economic, and environmental determinants of health. Community-based public health initiatives play a crucial role in this endeavour by leveraging local resources, partnerships, and cultural insights to develop targeted interventions that address the unique needs of diverse communities. This article examines the principles, strategies, and outcomes of community-based public health initiatives in addressing health disparities and promoting health equity. Health disparities, characterized by differences in health outcomes among various population groups, persist as a profound challenge in public health worldwide. Despite advancements in medical science and healthcare delivery, these disparities continue to affect individuals across socioeconomic, racial, ethnic, and geographic divides, undermining the overarching goal of achieving health equity. Addressing health disparities requires comprehensive strategies that extend beyond traditional clinical interventions to tackle the social, economic, and environmental determinants of health. In this context, community-based public health initiatives emerge as pivotal mechanisms for reducing disparities and promoting health equity. Community-based public health initiatives represent a proactive approach to health promotion and disease prevention, leveraging local resources [2], knowledge, and partnerships to address the unique needs of diverse communities. By engaging stakeholders at the grassroots level, including community members, healthcare providers, government agencies, and non-profit organizations, these initiatives develop tailored interventions that target the underlying drivers of health disparities. From enhancing access to healthcare services to promoting healthy behaviours and addressing social determinants, community-based initiatives play a critical role in creating environments that support health and well-being for all. This research aims to explore the principles, strategies, and outcomes of community-based public health initiatives in addressing health disparities. Through a comprehensive review

© Copyright IT Medical Team | **This article is available in:** https://www.itmedicalteam.pl/health-science.html

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19828-y

An experimental study of game theory with various word embeddings for automatic extractive text summarization

Minni Jain[1] · **Rajni Jindal**[1] · **Amita Jain**[2]

Received: 21 September 2023 / Revised: 3 July 2024 / Accepted: 6 July 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
Text Summarization is a process to abridge a long-size document into condensed form by comprising all the prime information and central theme. Researchers proposed numerous approaches for automatic text summarization, but still state of art methods are not able to capture semantic information of input text and preserve the text cohesiveness properly. To incorporate these features the proposed novel hybrid approach integrates evolutionary game theory with word embedding models. Here each sentence acts as a player and the meaning (senses) of words of sentences form strategy for the respective games. The payoff is determined using word embedding similarity measures between words and senses of sentences. The proposed hybrid method is capable of having the sentences (in summary) from the sentences (in the document) that share a strong relationship. This idea preserves textual cohesiveness effectively by modelling the interdependence between the sentences. The proposed work is evaluated and compared with existing state-of-the-art approaches on DUC02 (Task 1) datasets using the ROUGE-1, ROUGE-2, ROUGE-SU4 and cohesion score metric. The experimental results show the proposed method outperforms the state-of-art approaches.

Keywords Natural Language Processing · Word Embeddings · Extractive Text Summarization · DUC Datasets

1 Introduction

With the rapid development of the Internet and information-communication technologies, the volume of documents on the Internet is growing exponentially. Human beings can't analyse and mine the useful information from this massive unstructured data

✉ Amita Jain
amita.jain@nsut.ac.in

Minni Jain
minnijain@dtu.ac.in

Rajni Jindal
rajnijindal@dce.ac.in

[1] Computer Science and Engineering, Delhi Technological University, Delhi, India

[2] Computer Science and Engineering, Netaji Subhas University of Technology, Delhi, India

within the desired time. Therefore, it is required to provide the succinct form of data without losing its significance. Automatic Text Summarization (ATS) is an effective technique for producing a concise variant of a text that preserves the pertinent information of the text [1]. It reduces the human efforts and helps the user to get the desired information easily.

In literature, text summarization approaches can be categorised in to *Extractive* and *Abstractive*. Extractive summarization is hinged on taking out salient units (important sentences) from the document(s) and restructuring these extracted units to produce the desired summary. Extractive text summarization can be assumed to provide natural summaries by not altering the original structure of the text [1]. Abstractive summarization, on the contrary, is an approach in which summaries are produced by either rephrasing the words of the text or using new words, keeping the essence of the original text intact [2]. In this paper, the primary focus is on extractive summarization to produce indicative and generic summaries.

Research on text summarization started with the use of techniques like term or phrase frequency [3], key phrases [4], the significance of each term [5] and latent semantic analysis of the terms or phrases [6] to identify the relevant sentence(s) from a document. As time passed, significant efforts were put in to improve these automatic summarization techniques. Today, many of the techniques have graph scoring methods as their base [7] such as the semantic similarity of phrases based on WordNet [8], use local and global importance to identify the relevant sentence(s) [9] and generation of consistent extractive summaries by ranking the sentences [10, 11]. Text summarization research point towards handling the summarization task with the help of neural networks [12–14] and metaheuristic approaches [15–20].

Despite a lot of progress in recent years in the field of text summarization, the problem of producing consonant narrative summaries has still not been completely solved. The main disadvantages of existing approaches for automatic text summarization systems are that 1) ignore the semantics of text as they only focus on statistical rules such as the position of a sentence, and frequency of a word rather than the semantic/sematic of text in documents. 2) ignores the cohesion i.e., the relationships of the contexts that the sentences have with their neighbouring sentences. The proposed approach is to introduce a novel hybrid method that uses game theory with word embeddings to capture the semantic information of text and preserve the cohesiveness of the final summary. The motivation comes from the fact that game theory has proven to have predictive powers in collective decision-making scenarios and can perform consistent labelling of the data [21]. It was first initiated by Von Neumann and Morgenstern in 1994 as a mathematical framework to model the fundamentals of decision-making in like situations. Game-theoretic frameworks have since been used in various ways to gain knowledge on the use of language and evolution [22]. In literature various approaches utilized WordNet to find out the relationship between words [23]. However, WordNet has very limited coverage of words and does not involve word context. Unlike traditional approaches based on WordNet, word embeddings provide semantic and contextual-based similarity and incorporate large and real-world words. This is the main reason to combine word embeddings with the game theory approach. The proposed method is the first attempt to use game theory with word embedding models to accomplish the task of.

The approach introduced in this work finds and discusses the importance of each sentence using its similarity with every other sentence. The word embeddings (Word-2Vec, GloVe, BERT, ELMo and FastText) based similarities are calculated, taking into

account the meanings of the words of the sentences. Each sentence acts as a player and the meaning of words form strategy for the respective games. The approach follows the idea that if the senses (meanings) of the words of a sentence are strongly bonded with the senses (meanings) of the words of another sentence, then the two sentences must be similar. This preserves the cohesiveness in the resultant summary. The idea is expressed through an example in a later section.

On the evaluation front, manual assessment is subjective, costly and not suitable for real-time data collection. The approach proposed in this work is tested on the datasets introduced by DUC (Document Understanding Conferences) [24]. The dataset used in this work is DUC 2002. ROUGE scores have been computed for the evaluation using the Rogue Evaluation Toolkit [25]. Results produced by the algorithm are significant and beat the baseline scores claimed by DUC. In comparison with the various DUC claimed top performing systems, our approach shows better results and proves to be flexible and scalable despite being unsupervised. The summarization process, in general, serves as a useful tool for a range of applications varying from document indexing [26], to question-answering systems [27] and aiding information retrieval systems for document classification [28].

Problem statement: The ignorance of text semantics and cohesion in text summarization leads to work in hybrid approaches for better performance.

The objective of the proposed approach is to introduce a novel hybrid method that uses game theory with word embeddings to capture the semantic information of text and preserve the cohesiveness of the final summary.

The key contributions of the proposed work are:

1. For the first time, the proposed hybrid approach captures semantic/contextual information of input sentences in the text and preserves cohesiveness in the final summary using game theory and various word embeddings.
2. This is the first approach that amalgamates word embeddings with game theory for automatic text summarization.
3. The proposed approach investigates various word embedding models such as Word2Vec, FastText, GloVe, BERT and ELMo with game theory on the DUC2002 dataset.
4. Experimental results reveal that on evaluation matrices viz. ROUGE-1, ROUGE-2, ROUGE-SU4 and cohesiveness scores, the proposed approach outperforms the state-of-art methods on the DUC2002 standard dataset.

Roadmap: In Sect. 2, we demonstrate past work akin to text summarization. Section 3 states the prerequisite knowledge required to understand the algorithm proposed in this text. The proposed algorithm is then described in Sect. 4. Section 5 illustrates the method through an example. Section 6 describes the experimental setup and the evaluation results. In the last Sect. 7 discusses the conclusion of the work.

2 Related work

Text summarization has been an active area of interest in the research community for a very long time. Most of the studies in this area focus on extractive summarization. This literature review encompasses high-quality research conducted between 1996 and 2023. Over the years, the foundation of natural language has shifted towards mathematical models.

Applications of statistical techniques include language modelling for automatic speech recognition using smoothed n-grams, part-of-speech tagging using hidden markov models, syntactic parsing using probabilistic grammars and word sense disambiguation. The recent robust graph-based methods have proven to be beneficial in areas of natural language processing like keyword extraction [29], text summarization using lexical centrality [5] and word clustering [30]. The search for an optimal solution to the summarization problem has led to an extensive exploration of both extractive and abstractive methods.

Early methods of extractive summarization were based on simple features like the frequency of the words of the sentence, the semantic nature of the sentences in a document and the key phrases indicating the importance of those sentences. Approaches that were explored involved the representation of the documents and their positions, lexical chains [31, 32] and cue and stigma words [4].

The machine learning approach for text summarization involved classifying the sentences of the text as summary and non-summary sentences. The model was trained using a reference dataset, learning the classification probabilities at the time of training. This approach includes both unsupervised clustering methods [25] as well as supervised learning methods such as classification-based techniques [33]. The supervised learning-based approaches [12, 14, 34] require labelled summaries at the training phase whereas the unsupervised systems [35, 36] mostly rank sentences using some heuristics and do not require manually labelled summaries for training.

A common method known as term frequency-inverse document frequency (tf-idf) [37] reflects on how important a word in a document is using the measure inverse document frequency or *idf* (formulated as $\log(N/n_i)$, where N denotes the count of the documents in a collection and n_i denotes the count of the documents comprising of the word i). The system CLASSY [38] used *tf-idf* scores to rank sentences and is currently the best-performing system on the dataset DUC 2004 in terms of *ROUGE-1* score.

Approaches like the cluster-based method [39] involved the semantic nature of a document (expressed in a natural language like English). The system MEAD [8] used document clustering to obtain salient words (highly similar to the topic) called centroids. These centroids further helped in the selection of sentences for summarization.

The methods, CLASSY and MEAD are similar to Google's *PageRank* algorithm [40] and Kleinberg's *HITS* algorithm [41]. However, the systems mentioned above proceed without considering the global frame of reference. Hence, an optimization method has been proposed based on Integer Linear Programming (ILP) which treats text summarization as a knapsack problem [42]. The authors also mention the use of the ILP optimization for this task.

Graph-theoretic approaches, including *TextRank* [43] and *LexRank* [5] introduced connections between the sentences (nodes) by an edge if the two sentences share common words. These approaches can visualize inter-document and intra-document similarities. Researchers mentioned the use of a graph-theoretic approach based on a weighted graphical representation of the documents obtained by topic modelling [10]. Another renowned classical approach, Latent Semantic Analysis uses Singular Value Decomposition (SVD) to find the important sentences in the document [44]. In KLSum [45], the summary is built using the criteria of Kullback-Lieber (KL) divergence. Biased centrality ranking called KP-centrality works by grouping semantically related sentences based on some known similarity measure. The size of the group that a sentence belongs to determines the rank of its importance [46]. The sentences' relationships are modelled as a graph by the graph-based ranking methods, which then use random walk algorithms to rate the sentences [43]. The unsupervised graph-based ranking is expanded upon by the CoRank model [35]. To assess

sentences collectively, it can take into account data from numerous viewpoints, such as relationships between words and sentences and sentence-sentence interactions.

Other methods include HMM [12] and CRF [47] based approaches, support vector regression-based methods [48], latent semantic analysis (LSA) and relevance measures. The authors mention the use of a greedy approach to evaluate sentences for summarization [49]. An algorithm based on rhetorical statuses is used for summarizing scientific articles [24]. More advanced techniques have also been introduced to measure the similarity of a sentence with the discourse by using synonyms and anaphoric relations. Researchers have also tried to integrate fuzzy logic and neural networks [12] to find a solution to the problem of text summarization. One specific area is covered by Parihar et. al. in which authors summarized the scientific papers based on citation and utilized the game theory approach [50]. A method based on topic modelling and word embeddings for the extractive summarization of single documents [51]. For Hindi text summarization, a combination of the Title feature, Sentence length, Sentence position, Numerical Data, Thematic word, Proper Noun and Term frequency & Inverse Sentence Frequency is used by the authors [52]. To target summarizing legal documents, authors present an automated optimal DL-based legal text summarization (ODL-LTS) approach [53]. In a work, authors provided brief overview of challenges, application area, benefits of graph interpretation, summarization, and visualization, while providing existing tools and datasets available for graph processing and learning [54].

In the literature, different methodologies have employed WordNet to ascertain relationships between words. Nevertheless, WordNet is constrained by its limited coverage of words and the absence of consideration for word context. In contrast to conventional approaches relying on WordNet, word embeddings offer semantic and context-driven similarity, encompassing a vast array of real-world words.

The existing literature also demonstrates a lack in capturing semantic nuances and maintaining cohesiveness in the summarized content.

3 Preliminaries

In this section, an introduction to evolutionary game theory, Nash equilibrium and word embedding models is presented, which are a pre-requisite to understanding the algorithm proposed in this work.

3.1 Evolutionary game theory

Evolutionary Game Theory was first brought to light by Smith and Price in 1973. Evolutionary game theory induces an *inductive learning* process wherein a set of entities play games with other entities (called neighbors) repeatedly. The players update their understanding of the shape of the game at each iteration. Further strategies are developed according to what has proven to be fruitful previously. The strategy space of a player i can be defined as a mixed strategy profile $x_i = (x_1, \ldots, x_m)$, where m is the count of pure strategies and x_h denotes the probability of player i choosing the hth pure strategy. In every iteration, a player updates its strategy space in accordance with the payoffs obtained as a result of the games. It then assigns greater probability to the strategy which has gained higher payoff and the process is continued until a point of equilibrium is achieved – Nash Equilibrium [21, 23].

3.1.1 Nash Equilibrium

In a non-cooperative game involving two or more players, Nash Equilibrium—so named after the mathematician John Forbes Nash Jr.—is a suggested solution where each player is assumed to be aware of the equilibrium strategies of the other players and no player stands to gain by changing only their own strategy [55, 56]. Each player has a plan they want to use, and according to game theory, no person can gain anything by changing their strategy while the others' stay the same. This combination of tactics and their associated rewards form a Nash equilibrium.

we have used the discrete time version of the replicator dynamics equation to conduct the experiments in this work:

$$x^h(t+1) = x^h(t)\frac{u\left(e^h, x\right)}{u(x,x)}\forall h \in S \tag{1}$$

where at each quantum of time t, the players update their strategy space observing the developing environment until equilibrium is attained. x^h denotes the probability that player i chooses its hth pure strategy and e^h is the expected payoff of a pure strategy e.

3.2 Word embedding models

As we all know computers do not understand text, they only understand numbers. Thus we need word embeddings to transform the text into a digital representation for computers to understand and perform the required calculations. Some of the famous word embedding techniques are: Word2vec—It was the first technique to predict embeddings of the words [57]. It uses a small neural network architecture to calculate the similarity between words in the corpus. It determines which words in the corpus occur more frequently with the target word and calculates the embeddings. Glove—It was introduced as an advancement to the word2vec algorithm. Because we compare each word with all the words in the corpus in word2vec, with a large corpus size our algorithm is very slow. In Glove, it focuses on the word co-occurrence and creates word embeddings by combining the embeddings of related words. FastText—It adopted its idea of generating word embeddings from the word2vec algorithm but rather than comparing the words in the corpus, it goes one level deeper and compares the part of words and characters. This allows us to create more generalized versions of embeddings. Elmo—It stands for "Embeddings from Language Model ". In this architecture, it uses the character level tokens and processes them in a bidirectional LSTM to make the word-level embeddings. The LSTM processes the tokens in a right-to-left and left-to-right approach. BERT—The idea of embeddings in BERT is the same as Elmo. But instead of using LSTM, it uses a deeply connected bidirectional architecture to have a more concentrated embedding of the words.

4 Proposed work: automatic text summarization using game theory and word embeddings

This section outlines the methodology for automatic extractive text summarization, employing game theory and word embeddings. In this framework, each sentence (player) in the text is denoted as $s \in T$, with strategies corresponding to different word senses of

the sentence. The initial strategy profile $S_i = \{1, \ldots, c\}$, for each player, is uniformly and normally distributed, ranging from 1 to c, where c represents the total number of strategies for that player. This distribution, indicating the likelihood of selecting specific strategies, undergoes iterative updates until it converges to the optimal distribution.

The games are conducted utilizing pay-off matrices and sentence-similarity matrices for pairs of sentences, detailed later in the text. Scores are assigned to sentences based on the outcomes of these games, leading to a ranking of sentences in descending order of importance. The algorithm's stepwise implementation is described in the subsequent sections.

Step 1: Extract a list of all the tokens (sentences) from the text (document) (please refer to Algorithm 1).
Step 2: Compute the sentence-similarity matrix using word embeddings, which contains pairwise similarities between all the sentences and depicts each sentences' interaction with every other sentence (neighbour) (please refer to Algorithm 2).
Step 3: Compute the sense-similarity matrix between each pair of senses extracted from each word of every sentence. This will further be required to extract the payoff matrices (please refer to Algorithm 3).
Step 4: Construct a strategy profile for each sentence and assign it a probability distribution i.e., distribute a set of probabilities over all the viable strategies (senses) for each sentence (player) (please refer to Algorithm 4).
Step 5: Start the games using the mathematical equations of replicator dynamics. Every player is required to play a game with every other player in each iteration. Continue the games until Nash Equilibrium is achieved (please refer to Algorithm 5).
Step 6: The strategy profile of each player is updated as a result of these games. Use these updated profiles to assign scores to the sentences (please refer to Algorithm 6).
Step 7: Rank the sentences according to these scores to form the desired summary. Figure 1 Shows the overall flow of the proposed method.

Algorithm 1 divides the given text into tokens. Each token consists of a sentence from the text. All such tokens are saved into a list. The list of sentences is denoted by *list* and the total number of sentences is given by n. Using *list*, we compute an $n \times n$ similarity matrix

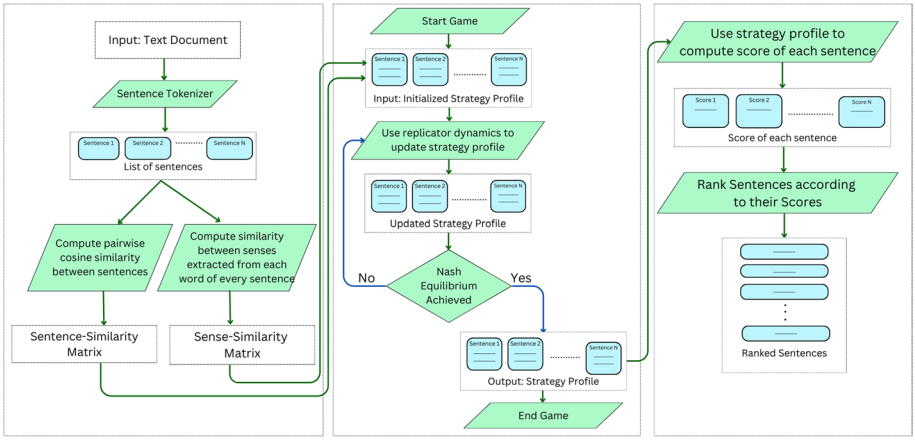


Fig. 1 Overall flow of the proposed text summarization method

– *sentence_matrix,* where every element w_{ij} denotes the similarity value between the sentences i and j. Proximity relations are taken into account by increasing the similarity value between two neighbouring sentences (can be checked using bi-gram) by a factor of the mean value of the *sentence_matrix*, as depicted in Algorithm 2.

Algorithm 1 Pre-Processing

Input: A text T
Output: A list of tokens (sentences) extracted from T – *list*
1. Input a text T
2. Extract from the text all the sentences into a *list*
3. Return the *list* of sentences

Algorithm 2 Sentence similarity matrix

Input: List *list* = (S_1, S_2, S_3, … , S_i, … , S_n) extracted from the text T
Output: A 2-D matrix (number_of_sentences × number_of_sentences) denoting the similarity

```
Let n = |list|, average = 0
Let sentence_matrix = 1 for all n * n values
for i = 1 to n do
    for j = 1 to n do
        Let score = 0, count = 0
        if (S_i is not equal to S_j)    // similarity between the same pair of sentences are ignored
            sentence1 = tokenize the sentence S_i and tag its words accordingly
            sentence2 = tokenize the sentence S_j and tag its words accordingly
            for each word1 in sentence1 do
                best_score = 0
                Let similarity_value_list = None
                for each word2 in sentence2 do
                    val = similarity-measure (word1, word2) using word embeddings
                    similarity_value_list. append(val)
                for k = 1 to | similarity_value_list | do   // filter out the 'None' values
                    if (similarity_value_list [k] is equal to 'None')
                        similarity_value_list [k] = 0.0
                best_score = Extract the maximum value from the similarity_value_list
                Add best_score to score
                Increment count by 1
            Divide score by count
            set sentence_matrix[i][j] = score
            add sentence_matrix[i][j] to average
Divide average by n * n
for i = 1 to n do
    for j = 1 to n do
        if (S_i is adjacent to S_j in list)
            add average to sentence_matrix[i][j]
Return sentence_matrix
```

The similarity value between a pair of sentences is determined through the following approach: for each word in a sentence, we compute its similarity with every word in the other sentence, employing diverse word embeddings such as word2vec, gloVe, fastText,

ELMo, and BERT. The highest obtained similarity value for each word pair is recorded. The average of these maximum values yields the overall similarity value between the two sentences. The comparison of these embeddings within the context of game theory for text summarization is elaborated upon in the results section (refer to Sect. 6).

The most popular, cosine similarity is used in Word2Vec word embedding models. Using this paradigm, the cosine similarity between the vectors of two words can be used to determine how similar the two words are. The cosine angles of the two vectors are multiplied to determine the cosine similarity.

$$\cos\alpha = \frac{AXB}{|A|X|B|} = \frac{\sum_{i=1}^{n} A_i X B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} X \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{2}$$

where A_i and B_i, respectively, are the elements of vectors A and B.

A list of unique senses (synsets) for every word of the text is extracted using a lexical database (WordNet 3.0 in this case) of the corresponding natural language. Let c denote the total number of available senses corresponding to all the words of the text. The payoff matrices are extracted from a $c \times c$ similarity matrix – *sense_matrix* constructed using the pairwise similarities of all the available senses. The reference knowledge base used in this work is WordNet 3.0 [58].

For each synset, we calculate its similarity with all other synsets. Again, to calculate the similarity measure between senses of words above mentioned word embeddings are used. Algorithm 3 gives a more computational and mathematical view for the same.

Algorithm 3 Sense Similarity Matrix

Input: List *list* = (S_1, S_2, S_3, ... , S_i, ... , S_n) extracted from the text *T*
Output: A 2-D matrix (number_of_synsets × number_of_synsets) denoting the pairwise similarities between all the synsets of all the words of the text – *sense_matrix*

```
Let all_synsets = None
for each sentence S_i in list do
    Tokenize the sentence and tag its words accordingly
    W = Extract all the words from sentence that are present in a lexical database W = (w_1, w_2, w_3, ... , w_i, ... , w_n)
    for each word in W do
        Extract all the unique synsets corresponding to word from the same lexical database – word-synsets
        for synset in word-synsets do
            all_synsets.append (synset)
Let len = |all_synsets|, sense_matrix = 0 for all len * len values
// all_synsets = (synset_1, synset_2, synset_3, ... , synset_i, ... , synset_j, ... , synset_len)
for i = 1 to len do
    for j = 1 to len do
        Set sense_matrix[i][j] = similarity-measure (synset_i, synset_j)
Return sense_matrix
```

The partial payoff matrix – Z_{ij} for each game between the players i and j can be obtained by extracting a sub-matrix from the sense similarity matrix (*sense_matrix*) corresponding to the entries corresponding to the senses of the two players. The $m_i \times m_j$ sub-matrix, thus obtained, is called the *payoff_matrix* for the game between players i and j, where m_i and m_j are the count of the senses of players i and j respectively. Greater the similarity value between the senses of the two words, greater is the motivation for the word to take that sense, and hence, utilize the strategy linked to that sense.

The more detailed explanation of how the payoff is determined using word embedding similarity measures between words and senses of sentences, emphasises the distinctiveness and effectiveness of this measure in capturing semantic relationships. The payoff determination in our approach is intricately tied to the semantic relationships between words and senses of sentences. We leverage word embedding similarity measures, utilizing embeddings from various models such as word2vec, gloVe, fastText, ELMo, and BERT. For each word in a sentence, we calculate its similarity with every word in the other sentence across these embeddings. This process is conducted comprehensively, considering all possible word pairs.

The uniqueness of our approach lies in its ability to capture semantic relationships. Word embeddings encode semantic information, and by assessing similarity across different models, we attain a more comprehensive understanding of the semantic space. The highest similarity value for each word pair is recorded, reflecting the strength of their semantic connection.

Aggregating these maximum similarity values and calculating their average yields the overall similarity measure between two sentences. This approach offers an effective means of determining payoff by considering the rich semantic associations embedded in various word embeddings. It allows us to encapsulate the intricacies of word meanings and sentence senses, contributing to a more robust and context-aware summarization process.

The sense inventories of the words of the text are used to create the strategy space for the games. A list $C = (1,..., c)$ comprising of all the unique senses is created marking the total space for the games. This strategy space S for the games can be defined in the following manner:

$$\begin{matrix} si1 & si2 & \ldots & sic \\ . & . & & . \\ . & . & \ldots & . \\ . & . & & . \\ sn1 & sn2 & \ldots & snc \end{matrix}$$

Each row corresponds to the strategy space of one particular player and each column corresponds to one particular sense. The components s_{ih} denote the probability with which a player i decides to move ahead with its hth pure strategy over all its other strategies i.e. the probability that the sentence s_i (player) chooses the sense s_h (strategy). Algorithm 4 describes the same.

Algorithm 4 Strategy Space

Input: List $list = (S_1, S_2, S_3, \ldots, S_i, \ldots, S_n)$ extracted from the text T
Output: A 2-D matrix (number_of_words x number_of_synsets) denoting the strategy profile of each sentence against all the available senses – *strategy_space*

1. Let *strategy_space* = 0 for all n * *len* values
2. for $i = 1$ to n do
3. for $j = 1$ to *len* do // Assigning a Normal Distribution to our Strategy Space
4. $$strategy_space[i][j] = \begin{cases} |M_i|^{-1} & \text{if sense } j \text{ comes from the sentence } i \\ & \text{where } |M_i| \text{ is the total sense-count of sentence } i \\ 0 & \text{otherwise} \end{cases}$$
5. Return *strategy_space*

The initial distribution of the mixed strategy space can be done either uniformly (normal distribution as in unsupervised learning) or can extract information from a sense-tagged corpus with pre-labelled values (supervised learning).

Normal distribution follows the following approach:

$$s_{ij} = \begin{cases} |M_i|^{-1}, if\ senses_{ij} \in M_i \\ 0, otherwise \end{cases} \tag{3}$$

where $M_i = \{1,..., m_i\}$ is the sense inventories of each word in a text, where m_i is the number of senses associated with the word i. The supervised initialization is explained in (Tripodi & Pelillo, 2017).

The system dynamics can now be started. Algorithm 5 gives a better understanding of the mathematics involved in the process. In each iteration, a player plays a game with each of its neighbors one by one. The payoff for the hth strategy is calculated as:

Algorithm 5 Replicator Dynamics

Input: *list*, *sentence_matrix*, *sense_matrix*, *strategy_space*
Output*: strategy_space* gets updated

```
Let number_of_iterations = 100                    // Assumption
for i = 1 to number_of_iterations do
    for j = 1 to n do                             // Player
        for k = 1 to n do                         // Neighbour
            Let payoff_matrix = a (sense_count_of_player x sense_count_of_neighbour)
            submatrix extracted accordingly from the sense_matrix
            The payoff for each strategy (current_payoff) is calculated according to the following equation:
            sentence_matrix[j][k] * dot-product (payoff_matrix, strategy_space[k])
            The strategy_payoff is calculated as Σ current_payoff
        The player_payoff is calculated as Σ dot-product (current_payoff, strategy_space[j]^T)
        Let updation_values for a player j = strategy_payoff / player_payoff
        Update strategy_space[j] according to the obtained updation_values
```

$$u_i\left(e^h, x\right) = \sum_{j \epsilon N_i} \left(w_{ij} Z_{ij} x_j\right)_h \tag{4}$$

The player's payoff for the iteration is calculated as:

$$u_i(x) = \sum_{j \epsilon N_i} x_i^T \left(w_{ij} Z_{ij} x_j\right) \tag{5}$$

Using this method, we can weigh the impact that each sentence has on the options that a candidate sentence has to consider to select its own meaning. The obtained payoff for a sentence i takes into account its own similarity with the sentence $j - w_{ij}$; the similarities among the senses of its words and the words of the sentence $j - Z_{ij}$, and the sense predilection of sentence $j - x_j$. Strategies pertaining to higher payoffs emerge at every iteration as a result of the selection criterion. After the process terminates, each player chooses its sense according to these restrictions.

The system may run for a pre-defined number of iterations or until convergence, as achieved by the following equation of Nash Equilibrium:

$$\phi_i = arg \max_{h=1,\ldots,c} x_{ih} \tag{6}$$

The result of the replicator dynamics assigns to each word of each sentence the sense that the word is most likely to assume in the given context. The top few senses of each word are taken to form the overall meaning of the sentence.

Now that we have the updated strategy profiles of the players, we know the optimal senses corresponding to each sentence. For each sentence, we compare its top senses (strategies) with every sense (strategy) of the other sentence to find the similarity between the two. Using the mean of the obtained data, we assign each sentence a similarity score as depicted in Algorithm 6. The sentence which has the maximum similarity is given the highest score. A descending order of the scores gives the final ranking of the sentences for the summary.

Algorithm 6 Assignment of Scores

```
Input: List list = (S1, S2, S3, ... , Si, ... , Sn) extracted from the text T, strategy_space
Output: An output array scores containing the similarity scores of each sentence
    Let sentence_synsets = None
    for each sentence in T do
        important_senses = Extract a list of the top important senses (synsets) using the strategy_space matrix
        sentence_synsets.append (important_senses)
    Let scores = []
    for each sentence Si in T do
        val = 0
        for each sentence Sj in T do
            for each sense si in sentence_synsets[i] do
                for each sense sj in sentence_synsets[j] do
                    add similarity_measure (sense Si, sense Sj) to val
        val = val / (|sentence_synsets|*|sentence_synsets|)
        scores.append (val)
    Return scores
```

5 Illustration through an Example

This section discusses an example to illustrate the algorithm proposed in this work. The following document has been taken as input to the algorithm:

A smart home speaker was the need of the hour. I was looking for the same on the Internet when I stumbled upon the perfect recommendation—the Echo Dot 3rd generation from Amazon. I immediately decided to move forward with it. The package containing the product was delivered to me within two days of its purchase. After using the product, I realised that the new Echo Dot from Amazon is the best compact smart speaker in the market by a wide margin. The quality of sound is extraordinary. Alexa has become smarter than ever with reasonable introduction to smart home control. The microphone sensitivity has also been improved significantly. However, I do feel that over a period of time the outer fabric cover might get dusty and the new power port is not as universal as it could have been.

The document is a review of the "Echo Dot 3rd Generation" from Amazon. The following steps illustrate the procedure to be followed to rank the sentences in order to produce a summary.

Algorithm 1 requires to divide the input text into nine sentences – S0, S1, S2, S3, S4, S5, S6, S7 and S8. We then proceed to calculate the sentence similarity matrix as illustrated in *Algorithm* 2. The sentence similarity matrix obtained for this example is shown in Table 1.

Algorithm 3 states the procedure to calculate the sense similarity matrix using word embeddings (Word2Vec in this illustration). The output obtained after this step is depicted in Table 2.

After this step we initialize our strategy space according to *Algorithm* 4 and assign normal probabilities to all the strategies (senses) corresponding to each player (sentence). We then proceed to perform Replicator Dynamics (*Algorithm* 5) as we initiate the games between the sentences. The scores assigned to each strategy (sense) of every sentence is shown in Table 3.

We now have the important strategies (senses) corresponding to every sentence in the text – Table 4. The score of every sentence is impacted by the very sense that is categorized as important for it. The scores thus assigned are shown in Table 5.

After the assignment of scores, a summary of size (the required number of sentences) '*k*' is expected to be generated. The top '*k*' sentences are then extracted on the basis of their scores and are given as the output. Let the required number of sentences for the summary ('*k*') be 5. Table 6 shows a summary of length five.

Table 1 Sentence similarity matrix

	0	1	2	3	4	5	6	7	8
0	0.000	0.588	0.333	0.291	0.554	0.310	0.336	0.265	0.335
1	0.541	0.000	0.624	0.254	0.540	0.259	0.166	0.228	0.258
2	0.333	0.624	0.000	0.624	0.333	0.333	0.333	0.333	0.333
3	0.268	0.275	0.532	0.000	0.685	0.270	0.164	0.254	0.320
4	0.392	0.510	0.305	0.619	0.000	0.541	0.167	0.218	0.352
5	0.447	0.444	0.333	0.430	0.735	0.000	0.560	0.400	0.555
6	0.334	0.178	0.261	0.185	0.187	0.506	0.000	0.571	0.375
7	0.350	0.356	0.305	0.358	0.358	0.353	0.639	0.000	0.785
8	0.351	0.340	0.333	0.374	0.419	0.369	0.315	0.675	0.000

Table 2 Sense Similarity Matrix *77 rows X 77 columns*

	0	1	2	3	4	5	6	…	75	76	77
0	1.000	0.235	0.117	0.133	0.285	0.461	0.307	…	0.133	0.285	0.190
1	0.235	1.000	0.111	0.125	0.266	0.285	0.285	…	0.125	0.266	0.272
2	0.117	0.111	1.000	0.333	0.133	0.142	0.266	…	0.375	0.133	0.090
3	0.133	0.125	0.333	1.000	0.153	0.166	0.307	…	0.470	0.153	0.100
4	0.285	0.266	0.133	0.153	1.000	0.363	0.363	…	0.153	0.500	0.210
5	0.461	0.287	0.142	0.166	0.363	1.000	0.400	…	0.166	0.363	0.222
6	0.307	0.285	0.266	0.307	0.363	0.400	1.000	…	0.307	0.363	0.222
7	0.235	0.333	0.111	0.125	0.266	0.285	0.285	…	0.125	0.266	0.454
8	0.111	0.105	0.315	0.400	0.125	0.133	0.250	…	0.588	0.125	0.086
9	0.153	0.142	0.142	0.153	0.181	0.200	0.200	…	0.166	0.181	0.111
10	0.250	0.235	0.117	0.133	0.285	0.307	0.461	…	0.133	0.285	0.190
…	…	…	…	…	…	…	…	…	…	…	…
74	0.285	0.266	0.133	0.153	0.500	0.363	0.363	…	0.153	0.666	0.210
75	0.000	0.000	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
76	0.153	0.142	0.142	0.153	0.181	0.200	0.200	…	0.166	0.181	0.111
77	0.000	0.000	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000

Table 3 Strategy Space (After Replicator Dynamics) *9 rows X 77 columns*

	0	**1**	**2**	**3**	**4**	**…**	**75**	**76**	**77**
0	9.881	9.881	4.940	0.494	1.000	…	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	0.000
8	0.000	0.000	0.000	0.000	0.000	…	0.000	0.000	9.881

Table 4 Important strategies (senses) corresponding to the sentences

0	1	2	3	4
need.n.01	hour.n.01	angstrom.n.01	smart.n.01	home.n.01
iodine.n.01	perfective.n.01	echo.n.01	point.n.09	coevals.n.01
iodine.n.01	angstrom.n.01	smart.n.01	home.n.01	speaker.n.01
package.n.01	two.n.0	days.n.01	purchase.n.01	angstrom.n.01
iodine.n.01	echo.n.01	point.n.09	market.n.01	margin.n. 01
quality.n.01	sound.n.01	angstrom.n.01	smart.n.01	home.n.01
control.n.01	introduction.n.01	smart.n.01	angstrom.n.01	smart.n.01
sensitivity.n.01	introduction.n.01	better.v.02	angstrom.n.01	smart.n.01
iodine.n.01	power.n.01	feel.n.01	time_period.n.01	time.n.01

Table 5 Score of the sentences

Sentence	S0	S1	S2	S3	S4	S5	S6	S7	S8
Score	2.490	2.505	2.311	2.481	2.535	2.567	2.566	2.520	2.640
Rank	(7)	(6)	(9)	(8)	(4)	(2)	(3)	(5)	(1)

Table 6 Final summary of the illustrative text

Rank	Sentences
1	However, I do feel that over a period of time the outer fabric cover might get dusty and the new power port is not as universal as it could have been
2	The quality of sound is extraordinary
3	Alexa has become smarter than ever with reasonable introduction to smart home control
4	After using the product, I realised that the new Echo Dot from Amazon is the best compact smart speaker in the market by a wide margin
5	The microphone sensitivity has also been improved significantly

6 Experimental setup, results and discussion

This section discusses datasets description, evolutionary metrics, and comparison with state-of-art methodologies.

6.1 Evaluation dataset description and metrices

We have conducted various experiments using the DUC 2002 dataset, which is part of the Document Understanding Conference (DUC). DUC 2002 comprises 567 news stories in the English language, organized into 59 clusters.

The ROUGE package was utilized in the suggested study to automatically evaluate the algorithm [21, 25]. DUC has assumed the toolset for evaluations of automatic summarization. A measure of the summary's quality can be found in the number of overlapping units, such as n-grams, word sequences, and word pairs between the reference summary and the candidate summary.

ROUGE-N is an *n*-gram recall measure and is computed as follows:

$$ROUGE - N = \frac{\sum_{S\in\{Ref\sum\}}\sum_{n-gram\in S}Count_{match}(n-gram)}{\sum_{S\in\{Ref\sum\}}\sum_{n-gram\in S}Count(n-gram)} \tag{7}$$

where n is the n-gram's size, Countmatch (n-gram) returns the greatest number of n-grams that appear simultaneously in a set of reference summaries and a candidate summary, and Count (n-gram) returns the number of n-grams that appear in the reference summaries.

The evaluation of summaries was previously done using a score based on recall in the older ROUGE packages. However, the evaluation is now done using three metrics: ROUGE-N Precision, ROUGE-N Recall, and ROUGE-N F-Score, where N can be any number between one and four. The ROUGE toolbox reports distinct scores for each of the following: 1, 2, 3, and 4 g as well as the SU4: skip bigram. The longest subsequence between generated and standard summaries is examined by ROUGE-L. The ROUGE Version 1.5.5 has been used for the evaluation of our system.

6.2 Comparison with DUC systems on DUC 2002 dataset

This section compares the proposed method with popular five DUC systems on DUC 2002 dataset. Brief description of these five-systems sys code 21, sys 27, sys 28, sys 29 and sys31 is given in Table 7. These systems are considered to have the best ROGUE scores and have been widely used for comparisons. Their summaries have been listed on the DUC-NIST

Table 7 A brief description of the top five systems participating in DUC 2002

System Code	System ID	System Description
21	Wpdv-xtr.v1	Supervised sentence classification with WPDV (Weighted Probability Distribution Voting)
27	Ntt.duc02	Supervised sentence classification with SVM (Support Vector Machines)
28	Ccsnsa.v2	Supervised sentence extraction with the Hidden Markov Model (HMM) and the LRM (Logistic Regression Model)
29	Kul.2002	Unsupervised sentence extraction + topic segmentation
31	ULeth 131 m	Unsupervised sentence extraction + text segmentation

website. A comparison-based evaluation of ROGUE-1, ROGUE- 2 and ROGUE-SU4 scores of proposed approach, the baseline and the top DUC mentioned systems is performed. The results are shown in Table 8 and Fig. 2. It is very clear that the proposed system based on word embedding and Game Theory (TSWE&GT) outperforms the others.

6.3 Performance comparison on DUC 2002 dataset

We have performed a comparative evaluation of our proposed system with widely used baseline approaches and state-of-the-art systems published in the literature on DUC 2002, as shown in Table 9.

Description of Baseline approaches are as follows:

TextRank [59] a graph-based algorithm depicts a document as a graph of sentences, with the edges connecting two sentences based on how similar they are.

LexRank [5] is to determine the significance of each sentence in the document, LexRank applies the idea of eigenvector centrality to the graph representation of sentences.

Table 8 Results on the DUC 2002 dataset (Stop-words included)

System	ROGUE-1	ROGUE-2	ROGUE-SU4
TSWE&GT (proposed)	**53.4**	**28.1**	**28.9**
System 27	50.1	24.1	25.3
System 31	49.1	23.8	25.1
System 28	48.3	23.2	24.7
System 21	47.9	22.8	24.1
DUC 2002 Baseline (Lead)	47.5	22.6	24.4
System 29	46.7	21.7	23.5

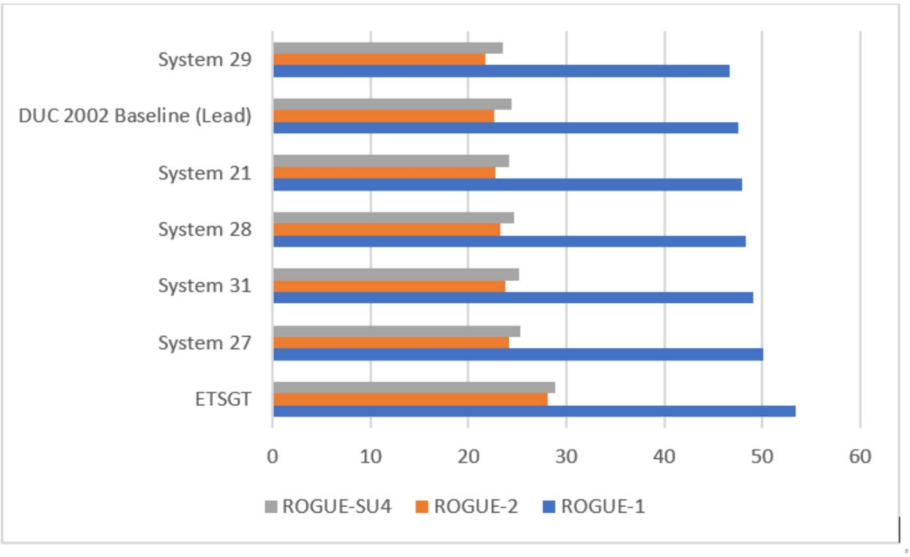


Fig. 2 Results for system comparison on DUC 2002 data (stop words included)

Table 9 ROUGE Scores on DUC 2002 using baseline, State-of-art and TSWE&GT approach

Methods	ROUGE-1	ROUGE-2	ROUGE-SU4
TextRank [59]	47.0	19.5	42.8
LexRank [5]	42.9	21.1	37.1
LSA [6]	43.0	21.3	40.0
KLSum [45]	38.3	16.9	32.2
SumBAsic [60]	39.6	17.3	35.1
ILP [61]	45.4	21.3	42.8
Egraph + coh [10]	47.9	23.8	-
Tgraph + coh [10]	48.1	24.3	-
URANK [9]	48.5	21.5	-
CoRank [35]	52.6	25.8	45.1
SummCoder [62]	51.7	27.5	44.6
KP-centrality [46]	49.8	24.5	43.7
GT_TS by Ahemed [63]	-	35.5	-
PE_MSC [51]	50.0	22.1	22.8
TSWE&GT (proposed)	53.4	28.1	46.7

LSA (Latent Semantic Analysis) [6]: By applying Singular Value Decomposition (SVD) on a document matrix D of size m x n, consisting of m sentences and n number of terms, LSA (Latent Semantic Analysis), attempts to identify salient sentences in the document.

KLSum [45]:The output summary of the document is produced by the text summarization algorithm KLSum based on the Kullback-Lieber (KL) divergence criterion.

SumBasic [60] is a greedy search approximation technique that combines a component to re-weight the word probabilities with a frequency-based phrase selection component to reduce redundancy.

Integer Linear Programming [61] A phrase-based summarization technique that uses integer linear programming to encapsulate the grammaticality restrictions across phrase dependencies.

The entity graph-based system Egraph [10] represents documents using a bipartite graph of phrase and entity nodes. They created a system called "Egraph + Coh." that has a coherence metric that is determined by utilising the average outdegree of an unweighted projection graph.

Tgraph [10] is a graph-based, unsupervised text summarising system that models topics using Latent Dirichlet Allocation (LDA). Parveen et al."Tgraph.'s + coh" proposal generates a summary by taking into account coherence determined using a weighted projection graph.

URANK [9] is a unified rank methodology built on a graph model that manages multi-document summarization at the same time.

The graph-based unsupervised extractive text summarization method CoRank [35] combines the graph-based ranking model with the word-sentence relationship.

SummerCoder [62] is an unsupervised framework for extractive text summarization based on deep auto-encoders.

GT_AS [63] is a game theory-based approach for text summarization. PE_MSC [51] is a text summarization approach using Entailment based minimum set cover.

The proposed approach demonstrated outstanding performance in extractive text summarization on the DUC 2002 dataset, as evidenced by impressive ROUGE scores, as

depicted in Fig. 3. Specifically, we achieved a ROUGE-1 score of 53.4, a ROUGE-2 score of 28.1, and a ROUGE-SU4 score of 46.7. When compared to baseline methods and state-of-the-art summarizers, including ILP, graph-based approaches like Tgraph, Egraph, URANK, and deep auto-encoder-based methods like SummerCoder, our proposed framework (TSWE&GT) surpassed them in terms of ROUGE-1, ROUGE-2, and ROUGE-SU4 scores on the DUC 2002 dataset. Table 10 presents the ROUGE scores for game theory hybridising with various word embedding models including Word2Vec, FastText, GloVe, BERT and ELMo on the DUC 2002 dataset. Figure 4 illustrates the graphical impact of these models. Notably, the ensemble Bert with game theory method exhibited superior performance compared to other models. Reason behind is that BERT considers both left and right context in its representations. It captures contextual information from the entire sentence. One limitation of this work lies in its reliance on the performance of embeddings for generating payoff matrices, a crucial aspect of the proposed text summarization approach. The effectiveness may be compromised when dealing with sentences containing metaphors, slang terms, and abbreviations, as the current approach does not address these linguistic elements.

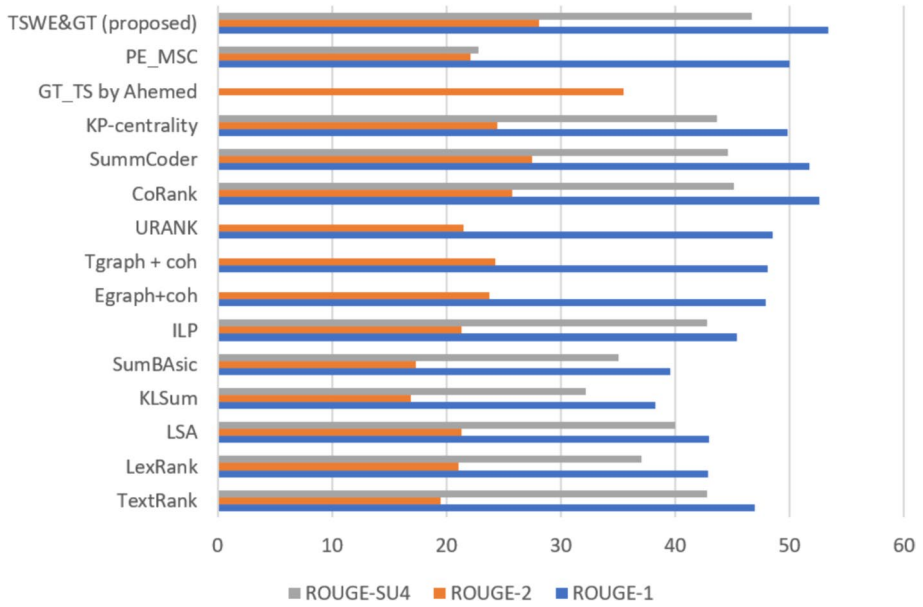


Fig. 3 ROUGE scores on DUC 2002 using baseline, state-of-the-art and TSWE&GT approach

Table 10 ROUGE scores for various word embedding models on DUC 2002 dataset

Word Embeddings	ROUGE-1	ROUGE-2	ROUGE-SU4
Word2Vec + GT	52.8	26.5	26.3
GloVe + GT	51.9	25.7	25.7
FastText + GT	52.3	26.2	26.8
BERT + GT	**53.4**	**28.1**	**28.9**
ELMo + GT	53.0	27.4	27.2

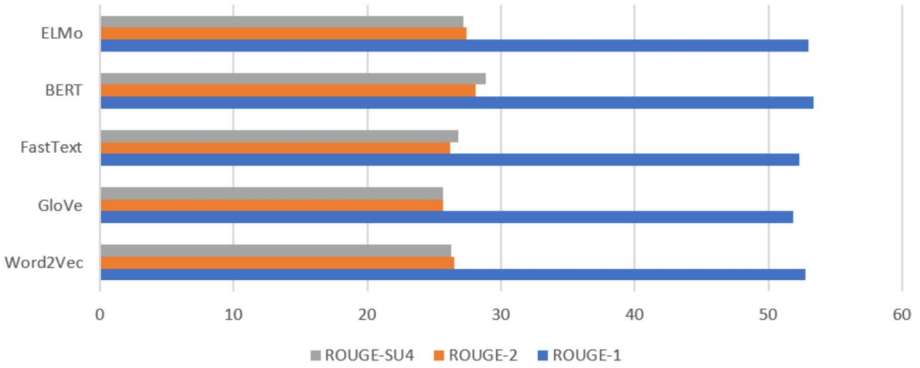


Fig. 4 Graphical Representation of results for word embedding models with game theory

6.4 Cohesiveness Assessment (S_{Coh})

This part compares the cohesion score of proposed methodology and above-mentioned state-of-art methods on DUC 2002 dataset. Cohesion is an essential element in text summarization. It refers to the degree to which sentences of resultant summary are connected for the smooth flow of reading and understanding. Cohesiveness in summary does not only mean grammatical correctness of sentences but it refers to the linking of the sentences at the both sentence level and passage level. The most used metric to compute cohesiveness of summary is cosine similarity.

$$SCoh = CosSim(Si, Sj) i \neq j = 1,2, \ldots k \tag{8}$$

where, S_i, S_j are the summary sentences. Higher the value of S_{Coh}, higher the connectivity between summary sentence and vice versa. Figure 5. shows the cohesion score of proposed method and baseline methods on DUC datasets (For simplicity four baseline methods with

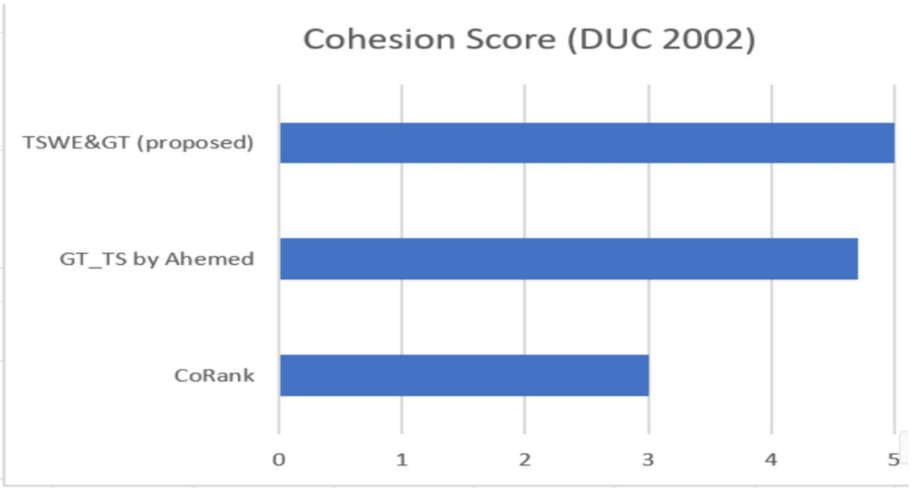


Fig. 5 Cohesiveness Assessment on DUC 2002 Dataset

highest rouge score are considered for cohesiveness comparison). It is observed from the analysis that proposed hybrid approach for text summarization using game theory and word embeddings has shown better cohesion value than the state-of-art for both datasets.

7 Conclusion and future work

A hybrid approach using evolutionary game theory and various word embedding models is capable to incorporate the semantic of input text and preserve the cohesiveness of the resultant summary. It is found that the proposed method successfully ranks each sentence in an order of importance by finding the mutual similarity of that sentence with all other sentences. The performance of generated summaries has been evaluated using ROUGE score on DUC 2002 dataset. Several experiments are performed and it has been found that proposed technique outperformed the state-of-art techniques on ROUGE-1, ROUGE-1 and ROUGE-SU4. Finally, cohesiveness is evaluated on DUC 2002 dataset and the result showed that proposed method preserves cohesiveness in final resultant summary.

An interesting future work direction can be the use of fuzzy game theory for text summarization. Hybridization of game theory with word embedding models can also enhance the performance of keyword extraction, query expansion, rumour detection etc. This work can be extended to cater to multi-document summarization with expectations of better results and greater efficiency.

Authors' contributions A.B. and B.C. conceived of the presented idea. A.C. developed the theory and performed the computations. A.B. and B.C. verified the analytical methods. B.C. encouraged A.B. to investigate and supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Funding No financial funding.

Availability of data and materials The generated/analysed data and code in the research work are available from the corresponding author on reasonable request.

Declarations

Ethical approval Not Applicable.

Conflicts of interests The authors declare that there is no financial/non-financial or personal nature conflict of interest.

References

1. Mutlu B, Sezer EA, Akcayol MA (2019) Multi-document extractive text summarization: A comparative assessment on features. Knowl-Based Syst. https://doi.org/10.1016/j.knosys.2019.07.019
2. Gupta S, Gupta SK (2019) Abstractive summarization: An overview of the state of the art. In Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2018.12.011
3. Luhn HP (2010) The Automatic Creation of Literature Abstracts. IBM J Res Dev. https://doi.org/10.1147/rd.22.0159
4. Edmundson HP (1969) New Methods in Automatic Extracting. Journal of the ACM (JACM). https://doi.org/10.1145/321510.321519
5. Erkan G, Radev DR (2004) LexRank: Graph-based lexical centrality as salience in text summarization. Journal of Artificial Intelligence Research. https://doi.org/10.1613/jair.1523

6. Steinberger J, Ježek K (2004) Text summarization and singular value decomposition. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-540-30198-1_25
7. Ferreira, R., De Souza Cabral, L., Lins, R. D., Pereira E Silva, G., Freitas, F., Cavalcanti, G. D. C., Lima, R., Simske, S. J., & Favaro, L. (2013) Assessing sentence scoring techniques for extractive text summarization. In Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2013.04.023
8. Sripada, S., Kasturi, V. G., & Parai, G. K. (2010) Multi-document extraction-based Summarization *. A Course Project.
9. Wan, X. (2010) Towards a unified approach to simultaneous single-document and multi-document summarizations. Coling 2010 - 23rd International Conference on Computational Linguistics, Proceedings of the Conference.
10. Parveen, D., Ramsl, H. M., & Strube, M. (2015) Topical coherence for graph-based extractive summarization. Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing. https://doi.org/10.18653/v1/d15-1226
11. Verma P, Om H (2019) A variable dimension optimization approach for text summarization. Adv Intell Syst Comput. https://doi.org/10.1007/978-981-13-0761-4_66
12. Cheng, J., & Lapata, M. (2016) Neural summarization by extracting sentences and words. 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers. https://doi.org/10.18653/v1/p16-1046
13. Kobayashi, H., Noguchi, M., & Yatsuka, T. (2015) Summarization based on embedding distributions. Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing. https://doi.org/10.18653/v1/d15-1232
14. Nallapati, R., Zhai, F., & Zhou, B. (2017). SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents. 31st AAAI Conference on Artificial Intelligence, AAAI 2017.
15. Abbasi-ghalehtaki, R., Khotanlou, H., & Esmaeilpour, M. (2016) Fuzzy evolutionary cellular learning automata model for text summarization.Swarm and Evolutionary Computation.https://doi.org/10.1016/j.swevo.2016.03.004
16. Alguliev, R. M., Aliguliyev, R. M., Hajirahimova, M. S., & Mehdiyev, C. A. (2011) MCMR: Maximum coverage and minimum redundant text summarization model.Expert Systems with Applications. https://doi.org/10.1016/j.eswa.2011.05.033
17. Alguliev RM, Aliguliyev RM, Isazade NR (2013) Multiple documents summarization based on evolutionary optimization algorithm. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2012.09.014
18. Aliguliyev RM (2009) A new sentence similarity measure and sentence based extractive technique for automatic text summarization. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2008.11.022
19. Rautray, R., & Balabantaray, R. C. (2017) Cat swarm optimization based evolutionary framework for multi document summarization. Physica A: Statistical Mechanics and Its Applications. https://doi.org/10.1016/j.physa.2017.02.056
20. Verma P, Om H (2019) Collaborative ranking-based text summarization using a metaheuristic approach. Adv Intell Syst Comput. https://doi.org/10.1007/978-981-13-1501-5_36
21. Tripodi R, Pelillo M (2017) A game-theoretic approach to word sense disambiguation. Comput Linguist. https://doi.org/10.1162/COLI_a_00274
22. Nowak MA, Komarova NL, Niyogi P (2001) Evolution of universal grammar. Science. https://doi.org/10.1126/science.291.5501.114
23. Jain, M., Suvarna, A. and Jain, A., (2022). An evolutionary game theory based approach for query expansion. Multimedia Tools and Applications, pp.1–25.
24. Simone Teufel, & Marc Moens. (2002) Summarizing Scientific Articles - Experiments with Relevance and Rhetorical Status. Computational Linguistics.
25. Nomoto T, Matsumoto Y (2001) A new approach to unsupervised text summarization. SIGIR Forum (ACM Special Interest Group on Information Retrieval). https://doi.org/10.1145/383952.383956
26. Pawar, D. D., Bewoor, M. S., & Patil, S. H. (2013) Context Based Indexing in Text Summarization Using Lexical Association. International Journal of Engineering Research & Technology. https://www.ijert.org/research/context-based-indexing-in-text-summarization-using-lexical-association-IJERTV2IS120142.pdf.
27. Arumae, K., & Liu, F. (2019) Guiding Extractive Summarization with Question-Answering Rewards. https://doi.org/10.18653/v1/n19-1264
28. Kallimani, J. S., Srinivasa, K. G., & Reddy B, E. (2010) Information retrieval by text summarization for an Indian regional language. Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering, NLP-KE 2010. https://doi.org/10.1109/NLPKE.2010.5587764

29. Biswas SK, Bordoloi M, Shreya J (2018) A graph-based keyword extraction model using collective node weight. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2017.12.025
30. Matsuo, Y., Sakaki, T., Uchiyama, K., & Ishizuka, M. (2006) Graph-based word clustering using a web search engine. COLING/ACL 2006 - EMNLP 2006: 2006 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference. https://doi.org/10.3115/1610075.1610150
31. Barzilay R, Elhadad M (1997) Using lexical chains for text summarization. Proceedings of the ACL Workshop on Intelligent Scalable Text Summarization. https://doi.org/10.3115/1034678.1034760
32. Silber HG, McCoy KF (2000) Efficient text summarization using lexical chains. International Conference on Intelligent User Interfaces, Proceedings IUI. https://doi.org/10.3115/1118253.1118294
33. Amini MR, Gallinari P (2002) The use of unlabeled data to improve supervised learning for text summarization. SIGIR Forum (ACM Special Interest Group on Information Retrieval). https://doi.org/10.1145/564396.564397
34. Begum, N., Fattah, M. A., & Ren, F. (2009) Automatic text summarization using support vector machine. International Journal of Innovative Computing, Information and Control.
35. Fang C, Mu D, Deng Z, Wu Z (2017) Word-sentence co-ranking for automatic extractive text summarization. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2016.12.021
36. Wang, D., Li, T., Zhu, S., & Ding, C. (2008) Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. ACM SIGIR 2008 - 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Proceedings. https://doi.org/10.1145/1390334.1390387
37. Christian, H., Agus, M. P., & Suhartono, D. (2016) Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF).ComTech: Computer, Mathematics and Engineering Applications. https://doi.org/10.21512/comtech.v7i4.3746
38. Conroy, J. M., Schlesinger, J. D., Goldstein, J., & O'Leary, D. P. (2004) Left-Brain/Right-Brain Multi-Document Summarization. Proceedings of DUC 2004, 4th Document Understanding Conference.
39. Aries, A., Zegour, D. E., & Hidouci, K. W. (2015).All Summarizer system at multiLing 2015: Multilingual single and multi-document summarization. SIGDIAL 2015 - 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference. https://doi.org/10.18653/v1/w15-4634
40. Page, L., Brin, S., Motwani, R., & Winograd, T. (1998) The PageRank Citation Ranking: Bringing Order to the Web. World Wide Web Internet And Web Information Systems. https://doi.org/10.1.1.31.1768
41. Kleinberg JM (2011) Authoritative sources in a hyperlinked environment. In The Structure and Dynamics of Networks. https://doi.org/10.1515/9781400841356.514
42. Hirao, T., Yoshida, Y., Nishino, M., Yasuda, N., & Nagata, M. (2013) Single-document summarization as a tree knapsack problem. EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference.
43. Berg-Kirkpatrick, T., Gillick, D., & Klein, D. (2011) Jointly learning to extract and compress. ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.
44. Steinberger, J., & Jezek, K. (2004) Using latent semantic analysis in text summarization and summary evaluation, Proceedings of the International Conference on Information System Implementation and Modelling (ISIM'04).
45. Haghighi, A., & Vanderwende, L. (2009) Exploring content models for multi-document summarization. NAACL HLT 2009 - Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Proceedings of the Conference. https://doi.org/10.3115/1620754.1620807
46. Wan X, Xiao J (2010) Exploiting neighborhood knowledge for summarization and keyphrase extraction. ACM Transactions on Information Systems doi 10(1145/1740592):1740596
47. Shen, D., Sun, J. T., Li, H., Yang, Q., & Chen, Z. (2007) Document summarization using conditional random fields. IJCAI International Joint Conference on Artificial Intelligence.
48. Galanis, D., Lampouras, G., & Androutsopoulos, I. (2012) Extractive multi-document summarization with integer linear programming and support vector regression. 24th International Conference on Computational Linguistics - Proceedings of COLING 2012: Technical Papers.
49. Carbonell J, Goldstein J (1998) Use of MMR, diversity-based reranking for reordering documents and producing summaries. SIGIR Forum (ACM Special Interest Group on Information Retrieval). https://doi.org/10.1145/3130348.3130369
50. Parihar, A.S., Jain, A. and Gupta, A., (2020), June. Citation-Based Scientific Paper Summarization Using Game Theory. In 2020 5th International Conference on Communication and Electronics Systems (ICCES) (pp. 1157–1161). IEEE.

51. Joshi A, Fidalgo E, Alegre E, Fernández-Robles L (2023) DeepSumm: Exploiting topic models and sequence to sequence networks for extractive text summarization. Expert Syst Appl 211:118442
52. Aote SS, Pimpalshende A, Potnurwar A, Lohi S (2023) Binary Particle Swarm Optimization with an improved genetic algorithm to solve multi-document text summarization problem of Hindi documents. Eng Appl Artif Intell 117:105575
53. Vaissnave V, Deepalakshmi P (2023) Modeling of automated glowworm swarm optimization based deep learning model for legal text summarization. Multimedia Tools and Applications 82(11):17175–17194
54. Mishra P, Kumar S, Chaube MK (2023) Graph Interpretation, Summarization and Visualization Techniques: A Review and Open Research Issues. Multimedia Tools and Applications 82(6):8729–8771
55. Choudhry, A., Khatri, I., Jain, M. and Vishwakarma, D.K., (2022). An Emotion-Aware Multitask Approach to Fake News and Rumor Detection Using Transfer Learning. IEEE Transactions on Computational Social Systems.
56. Jain, M., Jaswani, A., Mehra, A. and Mudgal, L., (2022). EDGly: detection of influential nodes using game theory. Multimedia Tools and Applications, pp.1–23.Egger, R. (2022) Text Representations and Word Embeddings: Vectorizing Textual Data." In Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications, pp. 335–361. Cham: Springer International Publishing.
57. Wu Z, Palmer M (1994). Verbs semantics and lexical selection. https://doi.org/10.3115/981732.981751
58. Mihalcea R, Tarau P (2004) TextRank: Bringing order into texts. Proceedings of EMNLP doi 10(3115/1219044):1219064
59. Nenkova A, McKeown K (2011) Automatic summarization. In Foundations and Trends in Information Retrieval. https://doi.org/10.1561/1500000015
60. Woodsend, K., & Lapata, M. (2011) Learning to simplify sentences with quasi-synchronous grammar and integer programming. EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference.
61. Joshi A, Fidalgo E, Alegre E, Fernández-Robles L (2019) SummCoder: An unsupervised framework for extractive text summarization based on deep auto-encoders. Expert Syst Appl 129:200–215
62. Ahmad A, Ahmad T (2019) A game theory approach for multi-document summarization. Arab J Sci Eng 44(4):3655–3667
63. Gupta A, Kaur M, Mittal S, Garg S (2021) PE-MSC: partial entailment-based minimum set cover for text summarization. Knowl Inf Syst 63(5):1045–1068

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

An Implementation of a Frequency Decomposition Algorithm for Power Quality Smoothening in a Hybrid SPVA-BES Microgrid

Divyank Srivastava, *Graduate Student Member, IEEE*, Vivek Narayanan, *Senior Member, IEEE*, Bhim Singh, *Fellow, IEEE*, and Arunima Verma, *Senior Member, IEEE*

***Abstract*—This article presents a novel application of Fourier decomposition method in a double-stage grid-tied microgrid integrating a solar photovoltaic array (SPVA) and battery energy storage (BES). Due to intermittent nature of SPVA generation and the nonlinearity of local loads connected at point of common coupling, the grid is subject to harmonic distortions. The proposed control method decomposes nonlinear and nonstationary signals to extract fundamental components necessary for generating switching pulses for voltage source converter. To ensure reliable operation during grid outages, system incorporates a BES, managed using a bidirectional dc–dc converter. This configuration maintains high-quality power delivery, and facilitates a smooth transition between grid-connected mode and standalone mode. Simulation and hardware prototype results validate effective operation of proposed control strategy.**

***Index Terms*—Adaptive control, battery energy storage (BES), frequency decomposition, IEEE Std. 519, power quality, solar photovoltaic array (SPVA).**

NOMENCLATURE

ANFIS	Adaptive neuro fuzzy interference system.
BES, BDDC	Battery energy storage, bidirectional dc–dc converter.
CEEMD	Complete ensemble empirical mode decomposition.
EMD, EEMD	Empirical mode decomposition, ensemble empirical mode decomposition.
FDM, FIBF	Fourier decomposition method, Fourier intrinsic band functions.
GCM	Grid-connected mode.
HTL, HCC	High to low, hysteresis current controller.
IRPT, InC	Instantaneous reactive power theory, incremental conductance.
LTH, LMS, LMS-LMF	Low to high, least mean square, least mean square-least mean fourth.
MG, MPP, MPPT	Microgrid, maximum power point, maximum power point tracking.
P&O, PSO, PR, PLL, PQ	Perturb and observe, particle swarm optimization, proportional resonant, phase locked loop, power quality.
RESs, ROGI	Renewable energy sources, reduced order generalized integrator.
SAM, SNR, SOGI, SPVA, SRFT, STS	Standalone mode, signal to noise ratio, second-order generalized integrator, solar photovoltaic array, synchronous reference frame theory, state transition switches.
THD, TOGI	Total harmonic distortion, third order generalized integrator.
VMD, VSC, VSSLMS	Variational mode decomposition, voltage source converter, variable step-size least mean square.
f_g	Grid frequency.
i_{ga}, i_{gb}, i_{gc}	Grid currents.
$i^*_{ga}, i^*_{gb}, i^*_{gc}$	Reference grid currents.
i_{La}, i_{Lb}, i_{Lc}	Load currents.
$i_{\mathrm{VSC}a}, i_{\mathrm{VSC}b}, i_{\mathrm{VSC}c}$	VSC currents.
$i_{Lae}, i_{Lbe}, i_{Lce}$	Extracted fundamental components of load currents.
I_{pv}, P_{pv}	SPVA current, SPVA power.
$K_{pd}, K_{id}, K_{pb}, K_{ib}$	Proportional and integral gains of PI controllers of BDDC.
$u_{gae}, u_{gbe}, u_{gce}$	In-phase positive sequence unit templates.
$u_{gqa}, u_{gqb}, u_{gqc}$	Quadrature positive sequence unit templates.
v_{ga}, v_{gb}, v_{gc}	Grid phase voltages.
$v_{gae}, v_{gbe}, v_{gce}$	Fundamental frequency extracted grid phase voltages from FDM control.
$v^{ref}_{La}, v^{ref}_{Lb}, v^{ref}_{Lc}$	Reference load voltages.
v_{La}, v_{Lb}, v_{Lc}	Load voltages.

Manuscript received 7 May 2023; revised 6 September 2023 and 21 May 2024; accepted 11 June 2024. This work was supported in part by the INDO-US under Grant RP03443, in part by FIST under Grant RP03195, and in part by SERB-NSC for awarding fellowships to the authors. Paper no. TII-23-1618. *(Corresponding author: Divyank Srivastava.)*

Divyank Srivastava and Arunima Verma are with the Electrical Engineering Department, Institute of Engineering & Technology, Lucknow 226021, India (e-mail: divyanksri.srivastava@gmail.com; arunima.verma@ietlucknow.ac.in).

Vivek Narayanan is with the Electrical and Computer Engineering Department, National University of Singapore, Singapore 119077 (e-mail: viveksw.narayanan7@gmail.com).

Bhim Singh is with the Electrical Engineering Department, Indian Institute of Technology, New Delhi 110016, India (e-mail: bhimsingh@iitdalumni.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2024.3423352.

Digital Object Identifier 10.1109/TII.2024.3423352

1551-3203 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

V_{term}	Amplitude of extracted PCC voltage.
V_L^{ref}	Amplitude of reference load voltages.
V_{dc}, V_{mpp}, $V_{\text{dc}ref}$	DC-link voltage, maximum power point voltage, reference dc-link voltage.
W_{Le}, W_{pvff}, W_{ge}	Average weight of fundamental load current, weight of SPVA feedforward component, grid weight.
ω_o	Frequency of reference load voltages.
θ_g, θ_L	Phase angle of grid voltage, phase angle of load voltage.

I. Introduction

THE climate change motivates to upgrade electricity industry and creates a low-carbon economy to support future economic growth [1]. Researchers focus on RES because of environmental effects and depleting nature of conventional energy sources. microgrids (MGs) are formed by integrating nonconventional energy sources, like solar photovoltaic array (SPVA), wind turbines, fuel cells, and electric vehicles with existing loads on distributed networks. SPVA has significant advantages over other nonconventional sources [2].

One of major issues addressed while integrating SPVA into utility grid is its intermittent nature, as it is climate-dependent and highly stochastic. As a result, the load demand during peak times may be greater than rated limit, thus resulting in destabilization or even an outage. Pump storage or diesel generators are often employed to reduce peak load demand during an imbalance. Nowadays, batteries replace these storage sources for peak load management as a temporary storage source because of advancements in battery technology [3]. When battery energy storage (BES) is linked directly to dc-link, it experiences issues with a second harmonic and unregulated battery current, causing degradation to its effective lifespan. BES is incorporated into MG using a bidirectional dc–dc converter (BDDC). It protects BES from second harmonic issue during imbalanced loads [4].

Two major categories of SPVA transformer-less designs in the literature are double-stage and single-stage topologies [5]. Even though single-stage topologies possess some merits over double-stage, such as compactness, lower component count, and cheaper, they experience much strain on primary dc–ac voltage source converter (VSC), hence reducing overall reliability [6].

A maximum power point tracking (MPPT) algorithm is implemented in SPVA to improve energy production by monitoring and modifying solar panel's operating point equal to maximum power point (MPP) [7]. MPPT controller collects SPVA output voltage and current, determines output's power, and compares it with operating power. The controller modifies voltage to maintain SPVA output at MPP and ensures that SPVA generates most electricity possible [8], [9]. To estimate MPP of SPVA, numerous researchers have put forth various sorts of methodologies. Due to their simplicity, extensively used methods are Perturb and observe (P&O) and InC. Alternatively, artificial intelligence strategies like particle swarm optimization and adaptive neuro fuzzy interference system are developed for MPPT operation of SPVA [10], [11].

As MGs are situated near distribution end, they can be integrated and controlled so that they can operate in grid-connected mode (GCM) and standalone mode (SAM). There should be proper synchronization control following the IEEE std. [12], [13], which enables the system to switch smoothly between GCM and SAM. Using ac–dc power electronic converters, equipment that uses dc power includes computers, microwaves, battery chargers, and lighting systems in commercial and residential areas are operated and they are highly nonlinear and inject harmonics into grid [14]. Various control algorithms for VSC have been proposed in the literature to ensure the injection of high-quality power into grid and smooth operation of MG. Table I provides a comparison of these methods. Conventional control techniques, such as synchronous reference frame theory (SRFT) and Instantaneous reactive power theory (IRPT), as discussed in [15] and [16], have been utilized for power quality improvement. However, these techniques exhibit limitations, including poor harmonic rejection capabilities, inability to reject dc-offsets, suboptimal dynamic response, and high complexity. To address these shortcomings, adaptive-based control techniques, such as least mean square (LMS), variable step-size least mean square (VSSLMS), and LMS-least mean fourth (LMF), have been introduced in [17] and [18]. Despite their advancements, these adaptive methods also face challenges, including the inability to reject dc-offsets, slow convergence rates, and difficulties in accurately determining the phase and frequency of load and grid voltages. For estimation of phase and rejection of dc-offset, some generalized integrator-based controls have been developed like second-order generalized integrator (SOGI), CSOGI, third order generalized integrator (TOGI), and reduced order generalized integrator (ROGI) [19], [20]. Using abovementioned controls for low signal to noise ratio (SNR) signals, fundamental frequency can be extracted from load currents and grid voltages, but noise and distortion are still present in signals. Hence, decomposition approach is employed to solve this issue. Several techniques are available for analysis of nonstationary and nonlinear signals. Most popular and significant method for these signals is empirical mode decomposition (EMD), but certain drawbacks have been acknowledged in [21] and [22]. Furthermore, ensemble empirical mode decomposition [23] and complete ensemble empirical mode decomposition [24] have been developed. However, modifications done on EMD do not fundamentally or completely check problem of mode mixing, aliasing, and end effects. Ning et al. [25] presented variational mode decomposition (VMD) technique based on generalization of Wiener filters. But, it suffers in decomposing nearly lined-up modes when SNR is varied. An adaptive nonstationary and nonlinear decomposition method, namely Fourier decomposition method (FDM), is presented by Singh et al. [22]. Using Fourier method and Hilbert transform, FDM decomposes signal into a small set of Fourier intrinsic band functions (FIBFs).

This work introduces a novel application of FDM on a double-stage grid-tied SPVA-BES hybrid MG. Using this, power quality issues are resolved in VSC, and proper synchronization is implemented with smooth mode transfer between GCM and SAM.

Key contributions to this work are outlined here as follows.

1) FDM is employed in MG to extract fundamental component of grid voltages, thereby achieving higher-quality grid currents. FDM addresses harmonic distortion challenges prevalent in microgrids with substantial penetration of SPVA. Unlike previous studies, this method enhances FDM for dynamic frequency response analysis, crucial for stabilizing microgrids with fluctuating renewable energy outputs.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
COMPARISON OF DIFFERENT CONTROL METHODS

Sl. No	Algorithms	Example	Nature of control	Load balancing	Harmonics rejection	DC offset rejection	Dynamic response	Frequency and phase estimation	For Low SNR	
									Elimination of mode mixing	Aliasing and end effect rejection
1.	Conventional control methods	SRFT, IRPT [15-16]	Non-adaptive	Good	Poor	NO	Poor (takes 3-4 cycles to settle)	NO, Requires PLL and FLL	NO	NO
2.	Adaptive filter-type control methods	LMS, LMS-LMF, VSSLMS [17-18]	Adaptive	Very good	Better	NO	Good (takes 2-3 cycles to settle)	NO, Requires PLL and FLL	NO	NO
3.	GI-based control methods	SOGI [19]	Non-adaptive	Excellent	Good (Improves with cascading)	NO	Best (takes 1-2 cycles to settle)	YES, but for frequency estimation, FLL is required	NO	NO
		CSOGI [19]				NO				
		TOGI [19]	Adaptive			YES				
		ROGI [20]				YES				
4.	Decomposition based control methods	EMD [21]	Adaptive	Excellent	Good (Improves with cascading)	YES	Best (takes 1-2 cycles to settle)	NO, Requires PLL and FLL	NO	NO
		VMD [25]				YES			NO	NO
		FDM [22]				YES			YES	YES

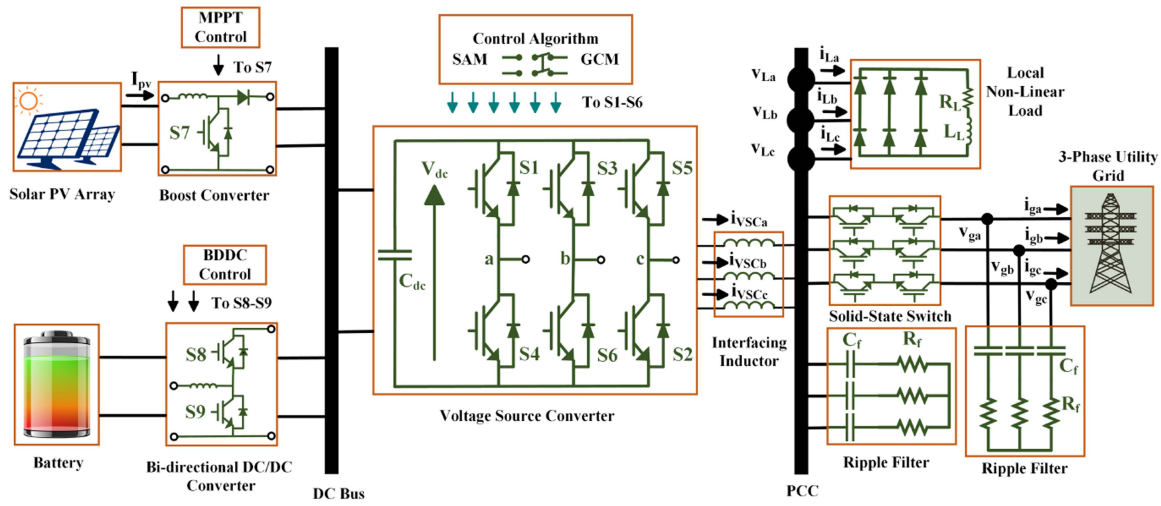


Fig. 1. System topology.

2) During transition between GCM and SAM, load currents experience significant transients, resulting in low SNR load currents. FDM technique decomposes load current into its fundamental components, noise, and harmonics. The fundamental component is then effectively extracted to maintain improved power quality. Furthermore, smooth synchronization between different operating modes is achieved. The authors have integrated FDM with control algorithms not extensively explored in existing literature, allowing for more precise and adaptive control of power flow, thereby improving the overall efficiency and reliability of microgrid operations.
3) FDM is refined to increase computational efficiency and accuracy. The authors have achieved this through parameter tuning, which reduces computation time while increasing the resolution of harmonic analysis. This refinement is critical for real-time grid management, ensuring the system operates efficiently and reliably.

II. SYSTEM LAYOUT

Fig. 1 presents system topology. It comprises an SPVA and BES. They are integrated to point of common coupling (PCC) using VSC, which transforms dc–ac conversion. SPVA is integrated at dc link using a dc–dc converter. BES is integrated using BDDC, which controls its charging and discharging currents. MG is integrated into the utility grid using STS, which are controlled to realize smooth synchronization and desynchronization of MG with the grid. To dampen ripples in PCC voltages, a parallel combination of resistor and capacitor is used at the grid and PCC. Nonlinear loads are composed of a diode bridge rectifier with resistive and inductive loads connected at PCC.

III. SYSTEM CONTROL DESIGN

MG consists of a distributed centralized controller that coordinates control and monitors overall operations of MG. Primary function of centralized controller is to determine if system can meet unanticipated load demand and to plan smooth transition between GCM and SAM. Overall control is subdivided into several specific controls: control of dc–dc boost converter, control of VSC during GCM, control of VSC during SAM, and control of BDDC.

A. Control of DC–DC Boost Converter

Due to intermittent nature of SPVA, its MPP is affected by changes in irradiance and temperature. As presented topology is a double-stage, a dc–dc boost converter is implemented for harvesting maximum power from SPVA. Fig. 2 demonstrates

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

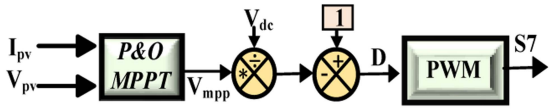


Fig. 2. DC–DC boost converter control.

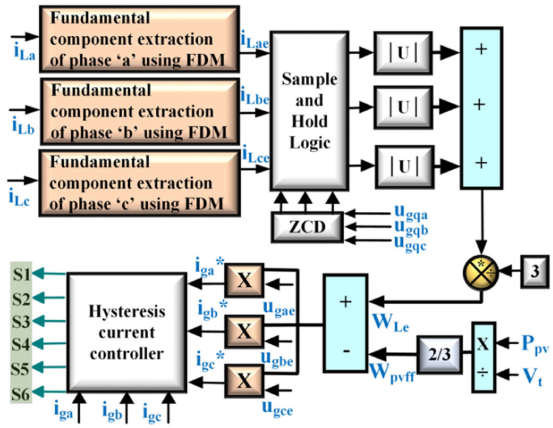


Fig. 3. Control of VSC in GCM.

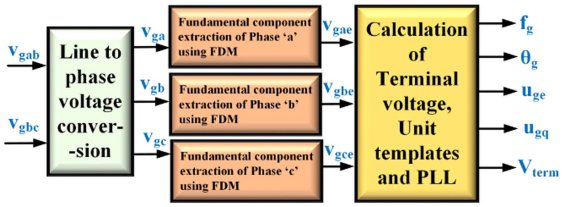


Fig. 4. Extraction of FIBFs of grid voltages.

control of dc–dc boost converter. For generating switching pulses of boost converter, a P&O MPPT algorithm is used, which gives MPP voltage (V_{mpp}), and using that, the duty cycle is generated for boost converter

$$D = 1 - \frac{V_{mpp}}{V_{\rm dc}}. \tag{1}$$

B. Control Strategy of VSC in GCM

In GCM, system encounters irregularities such as harmonics, voltage sags and swells, and nonlinearity in load currents. As depicted in Fig. 3, control strategy is designed to ensure delivery of high-quality power to grid while continuously supplying critical loads. Steps involved in GCM control are outlined as follows. By utilizing sensed line voltages (v_{gab}, v_{gbc}), grid phase voltages (v_{ga}, v_{gb}, v_{gc}) are determined

$$\begin{pmatrix} v_{ga} \\ v_{gb} \\ v_{gc} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ -1 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} v_{gab} \\ v_{gbc} \end{pmatrix}. \tag{2}$$

The application of FDM to grid voltages is utilized to extract fundamental components, as illustrated in Fig. 4. Using FDM, any data can be decomposed into a limited number of FIBFs. FDM provides a generalized Fourier expansion of a time series with varying amplitudes and frequencies. It is well-established that Fourier method is an ideal tool for time-frequency and spectrum analysis when processing signals. Fourier transform inherently provides an analytical representation of a signal, which, along with its Hilbert transform, elucidates its effectiveness.

As a result of this method, multicomponent and multifrequency signals are decomposed and extracted into fundamental and other FIBFs mono signals, which can be modeled as follows:

$$x(t) = \sum_{i=1}^{K} y_i(t) + N \tag{3}$$

where N represents noise or any residue components. $y_i(t)$ is K number of mono-component signals, which would be FIBFs in this framework. For FIBF $y_i(t)$ to proceed with nonstationary and nonlinear decomposition, it must meet requirements listed here.

1) FIBFs are zero mean function

$$\int_a^b y_i(t)dt = 0. \tag{4}$$

2) FIBFs are orthogonal functions

$$\int_a^b y_i(t)y_j(t)dt = 0; \quad i \neq j. \tag{5}$$

3) FIBFs provide analytic FIBFs with instantaneous frequency and amplitude always greater than zero

$$y_i(t) + jH(y_i(t)) = a_i(t)\exp(jj_i(t)) = z(t) \tag{6}$$

$$\text{where, } a_i(t), \frac{\partial}{\partial t}\phi_i(t) \geq 0 \,\forall t. \tag{7}$$

Followings are steps involved in decomposing a nonstationary, nonlinear signal using FDM.

Step 1: Use Fourier transform to get Fourier spectrum of signal $x(t)$.
Step 2: Utilizing spectrum of $x(t)$ acquired in Step 1 to express $z(t)$.
Step 3: Using LTH and HTL scan, obtain $z_k(t)$.
Step 4: From real part of $z_k(t)$ obtained in Step 3, obtain FIBFs $y_k(t)$.
Step 5: Reconstruct $x(t)$ as a summation of FIBFs $y_k(t)$ obtained in Step 4.

To generate FIBFs, two frequency scan techniques named LTH and HTL are used, as shown in Figs. 5 and 6.

Using FDM, fundamental grid voltage and load current can be extracted. Flowchart of FDM is shown in Fig. 7. Following extraction of the FIBFs of grid voltages (v_{ge}), amplitude of PCC voltages (V_{term}), along with in-phase (u_{ge}), and quadrature unit templates (u_{gq}), can be computed as outlined in [26].

Subsequently, SPVA feed-forward component and active weight of extracted fundamental load current are calculated to determine grid weight. SPVA feed-forward component is computed as follows:

$$W_{pvff}(k) = \frac{2}{3}\left\{\frac{P_{pv}(k)}{V_{\rm term}}\right\}. \tag{8}$$

For calculating active weight of fundamental extracted load currents, each current is processed with sample and hold block and then average weight is calculated as follows:

$$W_{Le}(k) = \frac{W_{Lae}(k) + W_{Lbe}(k) + W_{Lce}(k)}{3}. \tag{9}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

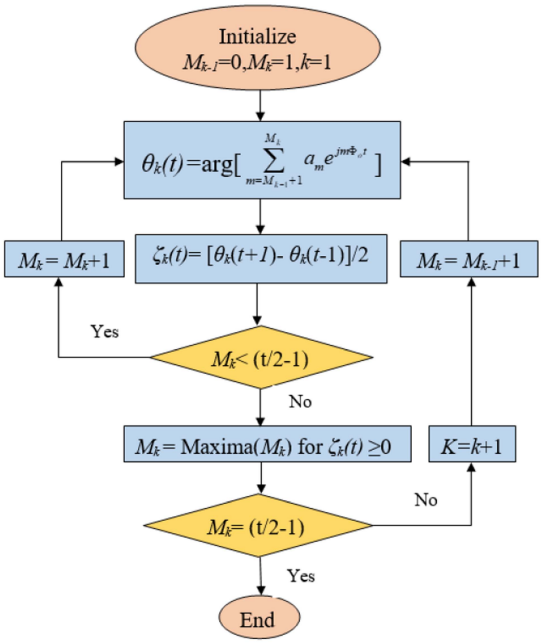


Fig. 5. Flowchart of LTH scan.

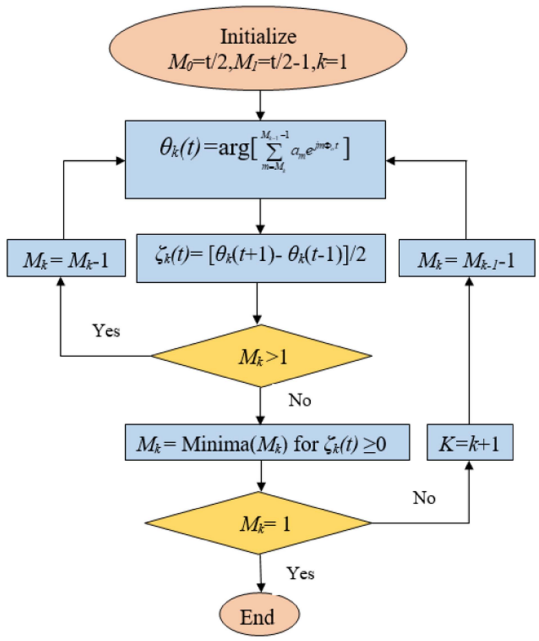


Fig. 6. Flowchart of HTL scan.

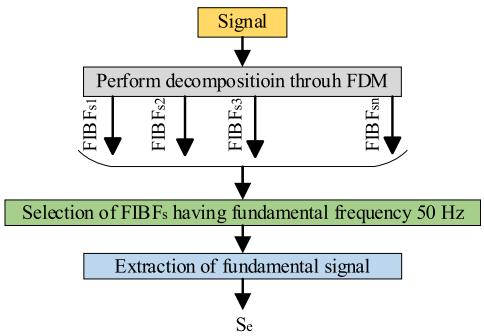


Fig. 7. Flowchart of FDM.

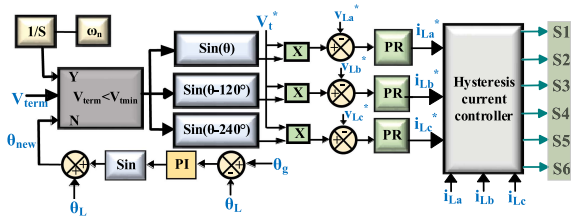


Fig. 8. VSC control during SAM.

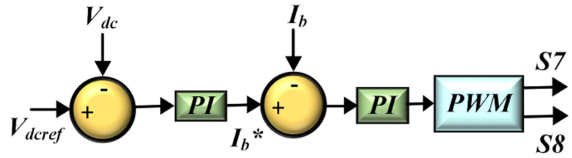


Fig. 9. BDDC control.

Grid current weight can be computed as

$$W_{ge}(k) = W_{Le}(k) - W_{pvff}(k). \tag{10}$$

For getting reference grid currents, grid currents' weight is multiplied with unit templates as follows:

$$i_{ga}{}^{*} = W_{ge} \times u_{gae}, i_{gb}{}^{*} = W_{ge} \times u_{gbe}, i_{gc}{}^{*} = W_{ge} \times u_{gce}. \tag{11}$$

To produce switching pulses of VSC in GCM, reference grid currents ($i_{ga}{}^{*}, i_{gb}{}^{*}, i_{gc}{}^{*}$) and sensed grid currents (i_{ga}, i_{gb}, i_{gc}) are given to hysteresis current controller (HCC).

C. Control Strategy of VSC in SAM

During SAM operation, primary objective of the VSC is to ensure uninterrupted power to load while maintaining high-quality load voltage in compliance with the IEEE-519 standards. Fig. 8 illustrates control mechanism of VSC during SAM. Reference load voltages are generated as follows:

$$v_{La}{}^{ref} = V_L{}^{ref} \sin(\omega_o t), v_{Lb}{}^{ref} = V_L{}^{ref} \sin(\omega_o t - 120^{\circ})$$
$$v_{Lc}{}^{ref} = V_L{}^{ref} \sin(\omega_o t - 240^{o}) \tag{12}$$

where $V_L{}^{ref}$ is desired magnitude of PCC voltages, and ω_o is equal to $2 \times \pi \times 50$ rad/s. When terminal voltage is greater than specified limit, then load phase angle is calculated as follows:

$$\theta_{\text{new}} = \theta_L + \sin\{k_p \times (\theta_g - \theta_L) + k_i \times (\theta_g - \theta_L)\}. \tag{13}$$

After computation of reference load voltages, these voltages are compared with sensed load voltages and then input into PR controller. PR controller subsequently generates reference currents for VSC. The advantage of using a PR controller over a PI controller is that it eliminates oscillations. Afterward, calculated reference VSC currents and actual VSC currents are sent to HCC to generate VSC switching pulses during SAM.

D. BDDC Control

Prime objective of BDDC is to control dc-link voltage. It also controls charging and discharging of battery during different dynamic conditions. Fig. 9 depicts BDDC control in which, first, $V_{\text{dc}ref}$ is compared with dc-link voltage, and it produces reference battery current. Afterward, resultant obtained by comparing reference battery current with sensed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

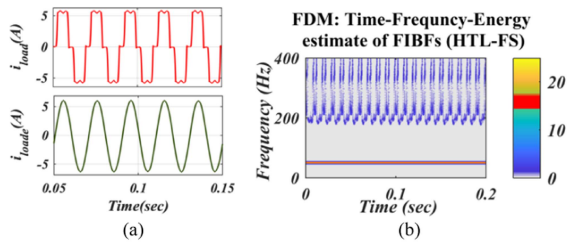


Fig. 10. (a) Extraction of load current using FDM. (b) TFEE of load currents.

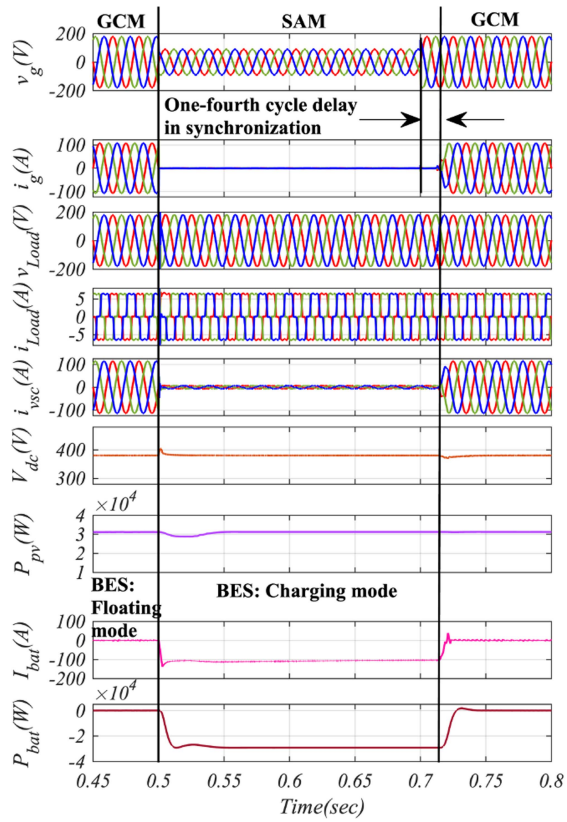


Fig. 11. Operation of the MG in a steady state during GCM and SAM, with smooth transition between these modes.

battery current is fed to PWM dc–dc block, which generates a switching pulse for S8 and S9 switches.

IV. Simulation Results

To illustrate dynamic function of control, performance of SPVA-interfaced MG is explored through simulation in MATLAB platform. Following modes show how well application of this FDM control on MG.

A. Operation of MG During GCM and SAM

Fig. 10(a) depicts nature of load current before and after extraction using FDM, respectively. These results clearly show that FDM is extracting pure fundamental load current, which is processed to generate reference grid current. Fig. 10(b) presents time-frequency energy estimate (TFEE) when extraction is done in HTL Fourier series, which indicates that fundamental component has highest energy, shown by red band and harmonics and noise having very less energy, shown from blue and gray color.

Fig. 11 illustrates the operation of MG in a steady-state during GCM and SAM, as well as smooth transition between these modes. It depicts utility grid voltages (v_g) and currents (i_g), load voltages ($v_{\rm Load}$) and currents ($i_{\rm Load}$), VSC currents ($i_{\rm vsc}$), dc-link voltage ($V_{\rm dc}$), SPVA power (P_{pv}), and battery current (I_{bat}) and power (P_{bat}). During both operational modes, MG functions smoothly, and THDs of load voltages and grid currents comply with IEEE 519 standard. In GCM, SPVA supplies power to the load, with any excess power injected into grid, while battery remains floating. In SAM, SPVA powers the load, and any surplus power is used to charge battery. At 0.5 s, a drop in grid voltage violates synchronization conditions, causing the system to seamlessly transition to SAM. At 0.7 s, although the grid voltage recovers, phase angle difference (θ_g - θ_L) does not meet the synchronization condition, so system continues to operate in SAM. After a one-fourth cycle delay, system transitions smoothly back to GCM.

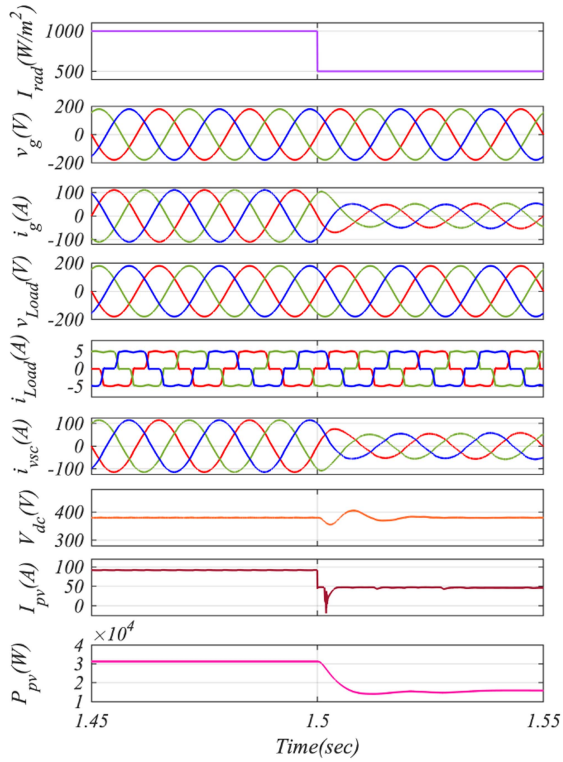


Fig. 12. Operation of MG under varying irradiance during GCM.

B. Operation of MG Under Varying Irradiance During GCM

Fig. 12 presents MG operation under varying irradiance on SPVA during GCM. Following results show solar irradiance (I_{rad}), grid voltages (v_g) and currents (i_g), load voltages ($v_{\rm Load}$) and currents ($i_{\rm Load}$), VSC currents ($i_{\rm vsc}$), dc-link voltage ($V_{\rm dc}$), and SPVA current (I_{pv}) and power (P_{pv}). At 1.5 s, the irradiance intensity decreases by 50%. The power produced by SPVA system decreases from 32 kW to 16.1 kW because of drop in irradiation. VSC performs multifunctional task of balancing load and maintaining dc-link voltage equal to $V_{{\rm dc}ref}$ using BDDC control. Grid currents waveforms are sinusoidal and adhere to the IEEE-519 standard.

C. Operation of MG During Load Unbalancing in GCM

Operation of MG under load unbalancing during GCM is shown in Fig. 13. It shows irradiation (I_{rad}), grid voltages (v_g) and currents (i_g), load voltages ($v_{\rm Load}$) and currents ($i_{\rm Load}$), VSC

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

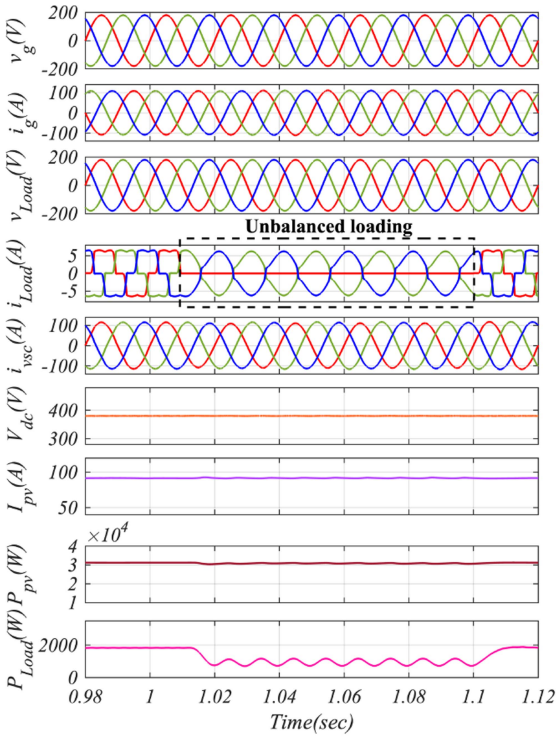


Fig. 13. Operation of MG under load unbalancing during GCM.

currents ($i_{\rm vsc}$), dc-link voltage ($V_{\rm dc}$), SPVA current (I_{pv}), SPVA power (P_{pv}), and load power ($P_{\rm Load}$). During removal of load in phase "a," loads connected to phases "b" and "c" function as a single-phase load. Reduction of phase "a" weight to zero leads to a decrease in fundamental active weight of load current. Despite these dynamic conditions, loads in the other two phases continue to operate smoothly, and grid currents are maintained as sinusoidal, in compliance with the IEEE-519 standard.

D. Operation of MG at Polluted Grid

Numerous MGs are connected to grid using power electronic converters, which consequently inject lower-order harmonics into grid. Additionally, harmonics are introduced into electrical system by nonlinear electronic switching components such as rectifiers, variable-speed drives, and electronic devices like computers, printers, TVs, servers, and telecom systems. The extensive connectivity of nonlinear loads to grid results in contamination of grid voltages by harmonics. Consequently, grid voltages at PCC are found to be polluted. Due to these polluted grid voltages, inaccurate unit templates and reference grid currents are generated. To address this issue, fundamental component of the grid voltages is extracted by processing polluted grid voltages using FDM, as illustrated in Fig. 14(a). Fig. 14(b) shows time-frequency energy estimate when extraction is performed in HTL scan, indicating that fundamental component has highest energy, represented by red band, while harmonics and noise exhibit significantly lower energy, represented by blue, light blue, and light green colors. Fig. 15 demonstrates extracted unit templates (u_{eabc}) and grid currents (i_{ge}) using extracted fundamental grid voltages, which are sinusoidal and comply with IEEE-519 standard.

V. Hardware Results

FDM control is tested on a prototype of a double-stage grid-connected system, developed in the laboratory, which validates

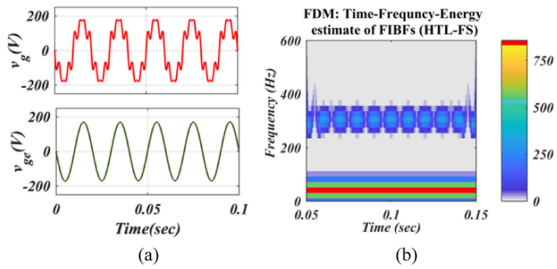


Fig. 14. (a) Extraction of grid voltages using FDM. (b) TFEE of grid voltage.

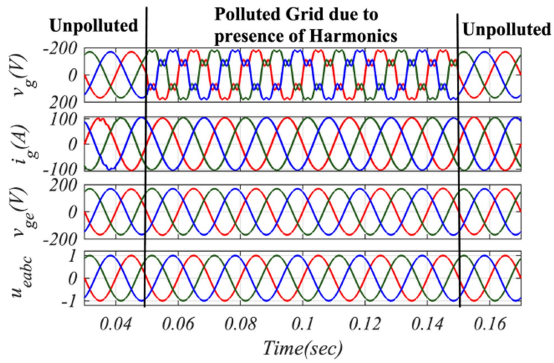


Fig. 15. Operation of MG with polluted grid.



Fig. 16. Hardware prototype.

its effectiveness and efficacy. A laboratory setup for MG prototype under investigation, used for experimental analysis, is shown in Fig. 16. Along with AMETAKE ETS 600/17 (SPVA), a battery bank is used to make a hybrid MG. Hall-effect based current (LEM LA55P) and voltage (LEM LV25P) sensors are used to detect signals. A dSPACE-1202 MicroLabBox DSP is used to deploy sensed signals to apply control. A power quality analyzer, HIOKI PQ3100, FLUKE-43B, and digital storage oscilloscope (MSOX3024T) are used.

A. Operation of MG in GCM at Steady-State

Fig. 17(a)–(f) presents operation of MG at a steady state in GCM. Fig. 17(a) and (b) presents waveforms of grid voltages and currents and analytical profile of grid voltages and currents. From analytical profile, it can be analyzed that grid voltage and grid currents THDs are 1.41% and 3.06%, which follows the IEEE 519 std. Fig. 17(c) presents the load voltages and currents waveform depicting that VSC successfully maintains uninterrupted power to critical load during GCM. Fig. 17(d)–(f) presents SPVA voltage and current, dc-link voltage, and SPVA

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

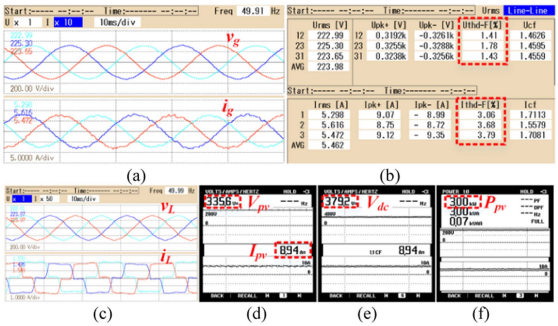


Fig. 17. Steady-state performance during GCM.

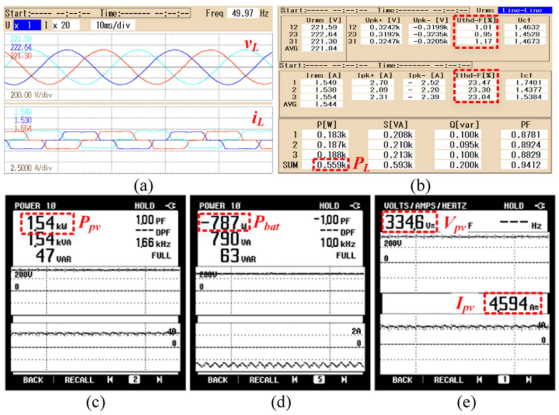


Fig. 18. Steady-state performance during SAM.

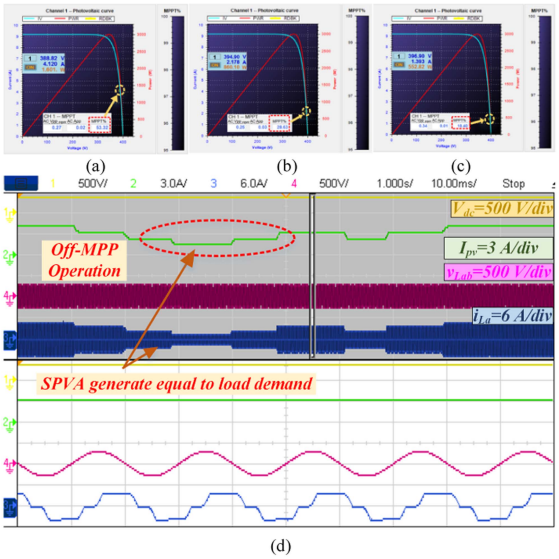


Fig. 19. Off-MPP operation of SPVA during SAM.

power. These show that dc-link voltage is regulated to desired voltage, and SPVA generates 3 kW power.

B. Operation of MG in SAM at Steady State

Fig. 18(a)–(e) presents operation of MG at steady state in SAM. Fig. 18(a) and (b) presents waveforms of load voltages and currents, load power, and harmonics spectra of load voltages and currents. Harmonics spectra show that PCC voltage THD is 1%, which follows the IEEE 519 standard. Fig. 18(c)–(e) presents SPVA voltage, current, and power (1.5 kW @500

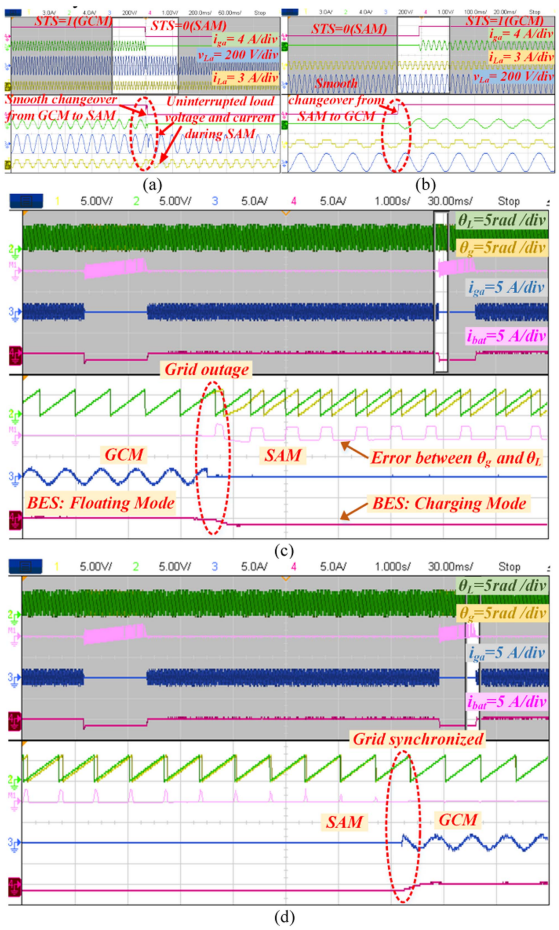


Fig. 20. Smooth changeover between GCM and SAM.

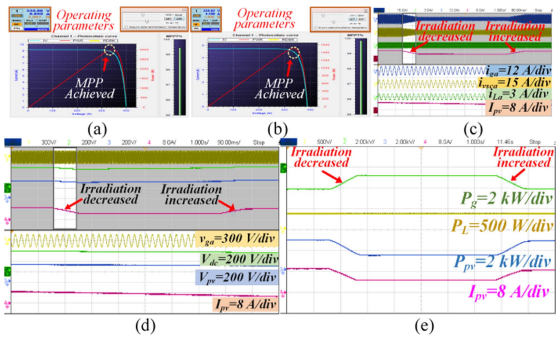


Fig. 21. Dynamic performance under varying solar irradiance.

W/m^2 irradiance), and power of battery. It shows that BES is charging. In the case of a fully charged battery during SAM, SPVA starts to work in OFF-MPP mode and generates power equal to load demand. Fig. 19(a)–(d) presents OFF-MPP operation. Fig. 19(a)–(c) shows OFF-MPP operating characteristics of SPVA at different loads. Fig. 19(d) represents dc-link voltage ($V_{\rm dc}$), SPVA current (I_{pv}), and load indices (v_{Lab}, i_{La}). It shows SPVA current change according to load currents, hence working in OFF-MPP mode.

C. Smooth Mode Changeover Between GCM and SAM

Fig. 20(a) and (b) shows an autonomous and seamless mode changeover of MG. It presents *STS*, i_{ga}, i_{La}, and v_{La}. It demonstrates how MG is safely separated from grid and runs in SAM during a power failure. The synchronization of MG

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

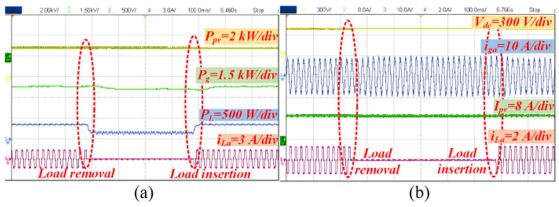


Fig. 22. (a) and (b) Dynamic performance under load unbalanced.

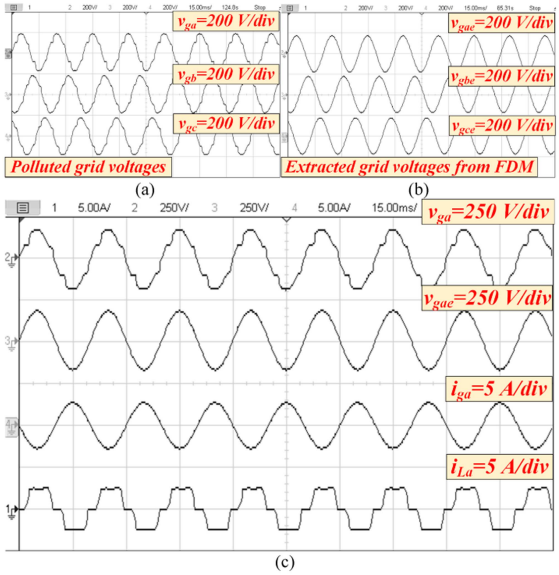


Fig. 23. Performance during polluted grid voltages.

into grid starts once grid voltages are recovered. When all synchronization requirements are met, MG is synchronized to grid by changing control pulse of STS to 1. Furthermore, it demonstrates that loads are being received in an uninterrupted power. Fig. 20(c) and (d) presents grid and load voltage phase angle (θ_g/θ_L), error between θ_g and θ_L, grid current (i_{ga}), and battery current (I_{bat}). It demonstrates that during transition from GCM to SAM, a phase angle mismatch occurs between θ_g and θ_L, and BES mode shifts from floating to charging. When grid is resynchronized, θ_g and θ_L realign, and system smoothly transitions back to GCM.

D. Operation of MG Under Varying Irradiance in GCM

Fig. 21(a)–(e) presents dynamic performance of this control under varying solar irradiance. Fig. 21(a) and (b) presents MPPT operation of SPVA at irradiation levels of 1000 W/m^2 and 500 W/m^2. Fig. 21(c) presents I_{pv}, i_{La}, $i_{\mathrm{vsc}\,a}$, and i_{ga}, from which it can be analyzed that when irradiance is decreased, SPVA current decreases, which results in decreased grid current and unchanged load current. Fig. 21(d) presents I_{pv}, V_{dc}, V_{pv}, and V_{ga}, from which it can be analyzed that when irradiance is decreased, all other three parameters remain unchanged. Fig. 21(e) presents I_{pv}, P_{pv}, P_L, and P_g, from which it can be analyzed that when irradiance is decreased, SPVA current decreases, due to which power produced by SPVA becomes half and power injected into grid also decreases. However, due to robust control, VSC is performing its objective of maintaining load power.

E. Operation of MG Under Load Unbalancing in GCM

Fig. 22(a) and (b) presents dynamic performance at load unbalancing. When load "a" is disconnected by a breaker due to a fault in phase "a," load current in that phase drops to zero. Consequently, remaining phases, "b" and "c," operate as a single-phase load, causing an increase in grid currents during this period. Fig. 22(a) presents i_{La}, P_L, P_g, and P_{pv}, illustrating that upon disconnection of phase "a" load, grid power increases, load power decreases, and the SPVA power remains unchanged. Fig. 22(b) presents i_{La}, i_{ga}, I_{pv}, and V_{dc}. It demonstrates that when phase "a" load is disconnected, SPVA current and dc-link

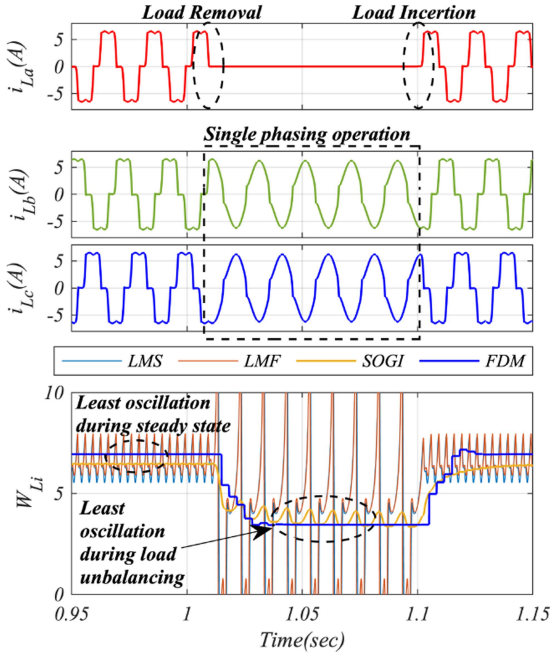


Fig. 24. Comparison of FDM with LMS, LMF, and SOGI controls.

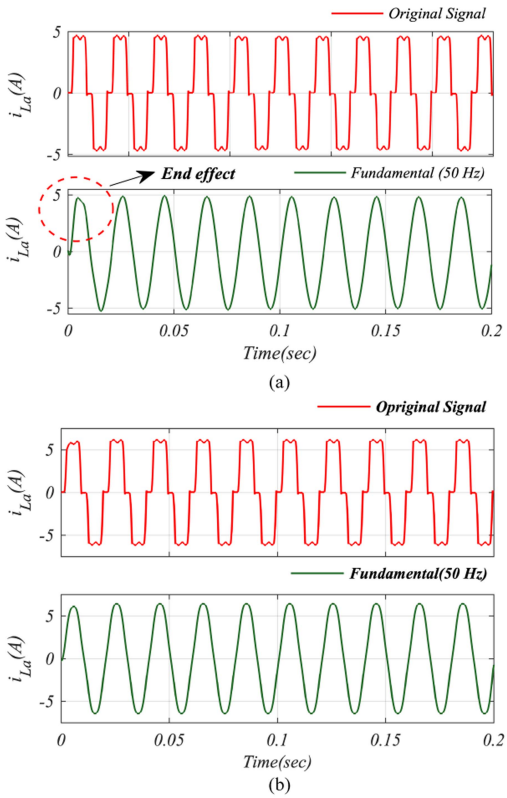


Fig. 25. Comparison of FDM with EMD.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

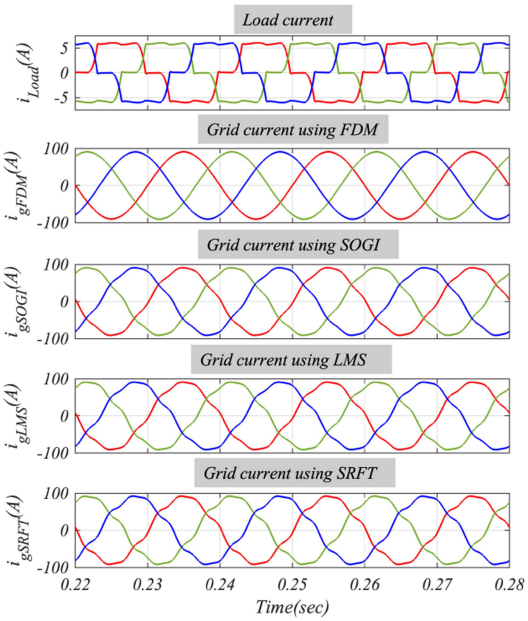


Fig. 26. Comparison of FDM with SOGI, LMS, and SRFT controls.

voltage remain unchanged. This indicates that SPVA continues to operate at its MPP regardless of operating conditions.

F. Operation of MG Under Polluted Grid Voltages

Fig. 23(a)–(c) presents experimental results under polluted grid conditions. Fig. 23(a) depicts polluted grid voltage waveforms for each phase. Fig. 23(b) shows extracted filtered grid voltage waveforms using an FDM filter, which is sinusoidal and complies with the IEEE-519 standard for THD. Fig. 23(c) presents polluted grid voltage (v_{ga}), extracted grid voltage (v_{gae}), generated grid current (i_{ga}), and load current (i_{La}). It is observed that generated grid current using filtered grid voltage is purely sinusoidal and adheres to the IEEE-519 standard, ensuring uninterrupted load supply.

G. Comparative Performance

A comparative performance of FDM over existing techniques during load unbalancing is shown in Fig. 24. It is observed that during steady-state, oscillation in the average load weight is least using FDM control as compared to LMS, LMF, and SOGI controls. Moreover, during load unbalancing, oscillation in average load weight is least using FDM control as compared to LMS, LMF, and SOGI controls.

The decomposition of load current using FDM and EMD is shown in Fig. 25. End effects are observed in EMD decomposition, while there is no end effect in FDM decomposition. Fig. 26 presents the grid current using FDM, SOGI, LMS, and SRFT controls. The grid currents using FDM are pure sinusoidal and free from harmonics, whereas harmonics are present in grid currents using SOGI, LMS, and SRFT controls.

VI. Conclusion

An SPVA-BES hybrid microgrid with a double-stage grid-connected system was developed in this work to demonstrate application of FDM technique. FDM technique eliminated effects of all harmonics and noises in grid voltages and load currents to estimate grid currents and provides fundamental sinusoidal components. The quality of extracted grid voltages, unit templates, and grid currents in GCM and load voltages in SAM follow the IEEE 519 standard. FDM technique presented efficient outcomes concerning performance in low SNR, oscillations, and convergence. BDDC control maintained dc-link voltage to desired value under all operating conditions; thus, it enhances reliability of the system. In SAM, a proportional resonant controller was used; hence, it reduces oscillations in load voltages. VSC performs multiobjective functions such as load balancing, smooth operation of system in both GCM and SAM, and seamless mode transition between grid-connected and standalone operations. Simulated responses of control technique were validated using a hardware prototype developed in the laboratory operating conditions.

Appendix

Parameters	Simulation	Hardware
Utility grid	220 V (L-L), 50 *Hz*	220 V (L-L), 50 *Hz*
SPVA	P_{mpp} = 32 *kW*, V_{mpp} = 342 *V*, I_{mpp} = 91.32 *A* (@1000 W/m^2)	P_{mpp} = 3 *kW*, V_{mpp} = 340 *V*, I_{mpp} = 8.7 *A* (@1000 W/m^2)
Interfacing inductor	L_f = 5 *mH*	L_f = 8 *mH*
Ripple filter	C = 10 μF and R = 5 Ω	C = 10 μF and R = 5 Ω
BES	V_{tb} = 240 *V*, 42 *Ah* K_{pd} = 0.6, K_{id} = 0.1, K_{pb} = 0.1, K_{ib} = 0.01	V_{tb} = 240 *V*, 42 *Ah* K_{pd} = 0.6, K_{id} = 0.1, K_{pb} = 0.1, K_{ib} = 0.01
Reference DC-link voltage and capacitance	V_{dcref} = 380 *V*, C_{dc} = 3 *mF*	V_{dcref} = 380 *V*, C_{dc} = 3 *mF*
Synchronization conditions	$-3° < \Delta\theta < 3°$, 49.5 *Hz* < f_g <50.5 *Hz*, 170 *V* < V_{term} < 190 *V*	$-3° < \Delta\theta < 3°$, 49.5 *Hz* < f_g <50.5 *Hz*, 170 *V* < V_{term} < 190 *V*

References

[1] Y. Zhang, L. Li, M. Sadiq, and F. S. Chien, "Impact of a sharing economy on sustainable development and energy efficiency: Evidence from the top ten Asian economies," *J. Innov. Knowl.*, vol. 8, pp. 1–9, 2023.

[2] B. Su, Z. Zhou, and H. Chen, "PVEL-AD: A large-scale open-world dataset for photovoltaic cell anomaly detection," *IEEE Trans. Ind. Inform.*, vol. 19, no. 1, pp. 404–413, Jan. 2023.

[3] R. Zafar and H. R. Pota, "Multi-timescale coordinated control with optimal network reconfiguration using battery storage system in smart distribution grids," *IEEE Trans. Sustain. Energy*, vol. 14, no. 4, pp. 2338–2350, Oct. 2023.

[4] A. C. Sunny and D. Debnath, "A novel low device count four-port converter based solar-fed off-grid system for catering household hybrid AC/DC loads," *IEEE Trans. Power Electron.*, vol. 38, no. 2, pp. 2658–2667, Feb. 2023.

[5] H. F. Ahmed, M. S. El Moursi, B. Zahawi, and K. Al Hosani, "Single-phase photovoltaic inverters with common-ground and wide buck–boost voltage operation," *IEEE Trans. Ind. Informat.*, vol. 17, no. 12, pp. 8275–8287, Dec. 2021.

[6] H. Mahdi, B. Hoff, and T. Ostrem, "A review of power converters for ships electrification," *IEEE Trans. Power Electron.*, vol. 38, no. 4, pp. 4680–4697, Apr. 2023.

[7] Z. Sun, Y. Jang, and S. Bae, "Optimized voltage search algorithm for fast global maximum power point tracking in photovoltaic systems," *IEEE Trans. Sustain. Energy*, vol. 14, no. 1, pp. 423–441, Jan. 2023.

[8] K. S. Tey, S. Mekhilef, M. Seyedmahmoudian, B. Horan, A. T. Oo, and A. Stojcevski, "Improved differential evolution-based MPPT algorithm using SEPIC for PV systems under partial shading conditions and load variation," *IEEE Trans. Ind. Inform.*, vol. 14, no. 10, pp. 4322–4333, Oct. 2018.

[9] D. Sera, L. Mathe, T. Kerekes, S. V. Spataru, and R. Teodorescu, "On the perturb and observe and incremental conductance MPPT methods for PV systems," *IEEE J. Photovolt.*, vol. 3, no. 3, pp. 1070–1078, Jul. 2013.

[10] K. K. Mohammed, S. Mekhilef, and S. Buyamin, "Improved rat swarm optimizer algorithm-based MPPT under partially shaded conditions and load variation for PV systems," *IEEE Trans. Sustain. Energy*, vol. 14, no. 3, pp. 1385–1396, Jul. 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[11] N. Priyadarshi, S. Padmanaban, J. B. Holm-Nielsen, F. Blaabjerg, and M. S. Bhaskar, "An experimental estimation of hybrid ANFIS–PSO-based MPPT for PV grid integration under fluctuating sun irradiance," *IEEE Syst. J.*, vol. 14, no. 1, pp. 1218–1229, Mar. 2020.
[12] *IEEE Application Guide for IEEE Standard1547, IEEE Standard for Interconnecting Distributed Resources with Electric Power Systems*, IEEE Standard 1547, 2-2008, 2009.
[13] *Photovoltaic (PV) Systems-Characteristics of the Utility Interface*, IEC Standard 61727 ed2.0, IEC, Geneva, Switzerland, 2004.
[14] P. Van Duijsen, J. Woudstra, and D. Zuidervliet, "Requirements on power electronics for converting kitchen appliances from AC to DC," in *Proc. Int. Conf. Domestic Use Energy*, 2019, pp. 190–197.
[15] S. F. Al-Gahtani, E. Z. M. Salem, S. M. Irshad, and H. Z. Azazi, "Improved instantaneous reactive power (PQ) theory-based control of DVR for compensating extreme sag and swell," *IEEE Access*, vol. 10, pp. 75186–75204, 2022.
[16] K. Santhosh, K. Chenchireddy, P. Vaishnavi, A. Greeshmanth, V. M. Kumar, and P. N. Reddy, "Time-domain control algorithms of DSTATCOM in a 3-phase, 3-wire distribution system," in *Proc. Int. Conf. Intell. Data Commun. Technol. Internet Things*, 2023, pp. 781–785.
[17] C. Balasundar, C. K. Sundarabalan, S. N. Santhanam, J. Sharma, and J. M. Guerrero, "Mixed step size normalized least mean fourth algorithm of DSTATCOM integrated electric vehicle charging station," *IEEE Trans. Ind. Inform.*, vol. 19, no. 6, pp. 7583–7591, Jun. 2023.
[18] M. H. Parveen, "A comparative study of control algorithms for DSTATCOM for harmonic elimination and reactive power compensation," *Int. Res. J. Eng. Technol.*, vol. 6, no. 3, pp. 3317–3330, Mar. 2019.
[19] Y. Zuo et al., "Active disturbance rejection controller for smooth speed control of electric drives using adaptive generalized integrator extended state observer," *IEEE Trans. Power Electron.*, vol. 38, no. 4, pp. 4323–4334, Apr. 2023.
[20] X. Yao, S. Huang, J. Wang, H. Ma, G. Zhang, and Y. Wang, "Improved ROGI-FLL-based sensorless model predictive current control with MRAS parameter identification for SPMSM drives," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 11, no. 2, pp. 1684–1695, Apr. 2023.
[21] X. Lang, Y. Zhang, L. Xie, P. Li, A. Horch, and H. Su, "Detrending and denoising of industrial oscillation data," *IEEE Trans. Ind. Informat.*, vol. 19, no. 4, pp. 5809–5820, Apr. 2023.
[22] R. P. Singh, S. D. Joshi, R. K. Patney, and K. Saha, "The Fourier decomposition method for nonlinear and nonstationary time series analysis," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 473, no. 2199, Mar. 2017, Art. no. 20160871.
[23] A. Caldeira and R. Coelho, "EEMD-IF based method for underwater noisy acoustic signals enhancement in time-domain," *IEEE Signal Process. Lett.*, vol. 30, pp. 294–298, Mar. 2023.
[24] D. Wang, Y. Y. Bao, and C. M. Wang, "A hybrid deep learning method based on CEEMDAN and attention mechanism for network traffic prediction," *IEEE Access*, vol. 11, pp. 39651–39663, 2023.
[25] Z. Ning, N. Cai, J. Zhao, W. Li, and H. Wang, "Error compensation for optical encoder based on variational mode decomposition with a coarse-to-fine selection scheme," *IEEE Trans. Instrum. Meas.*, vol. 72, Jan. 2023, Art. no. 1001510.
[26] B. Singh, A. Chandra, and K. Haddad, *Power Quality: Problems and Mitigation Techniques*. Hoboken, NJ, USA: Wiley, Jan. 2015.



Divyank Srivastava (Graduate Student Member, IEEE) was born in Lucknow, India, in 1995. He received the B.Tech. degree in electrical from the Institute of Engineering and Technology (IET), Lucknow, India, in 2016, and the M.Tech. degree in power systems from Delhi Technological University, New Delhi, India, in 2018. He is currently working toward the Ph.D. degree in electrical engineering from the IET, Lucknow, India.

His research interests include power system, power electronics, solar conversion systems, energy storage, power quality mitigation in distribution systems, and microgrids.



Vivek Narayanan (Senior Member, IEEE) was born in Kerala, India, in 1993. He received the B.Tech. degree in electrical and electronics engineering from the College of Engineering Thalassery, Kerala, India, in 2015, the M.Tech. degree in power electronics and drives from the Government College of Engineering Kannur, Kerala, India, in 2017, and the Ph.D. degree in power electronics and drives from the Department of Electrical Engineering, Indian Institute of Technology Delhi, New Delhi, India, in 2024.

He is currently a Postdoctoral Research Fellow with the National University of Singapore, Singapore. His research interests include power electronics, solar and wind energy conversion systems, energy storage, power quality improvement in distribution systems, and microgrids.

Dr. Narayanan was the recipient of the IEEE International Conference on Environment and Electrical Engineering /I&CPS Europe Best Poster Award and the IEEE ICPC2T Best Paper Award in 2019 and 2022.



Bhim Singh (Fellow, IEEE) received the B.E. degree in electrical from the University of Roorkee (Now Indian Institute of Technology Roorkee), Roorkee, India, in 1977, and the M.Tech. degree in power apparatus and systems and Ph.D. degree in electrical engineering from the Indian Institute of Technology Delhi (IIT Delhi), New India, in 1979 and 1983, respectively.

In 1983, he was with the Department of Electrical Engineering, University of Roorkee, as a Lecturer. He became a Reader there in 1988. In Dec. 1990, he was with the Department of Electrical Engineering, IIT Delhi, India, as an Assistant Professor, where he became an Associate Professor in 1994 and a Professor in 1997. He was ABB Chair Professor from Sep. 2007 to Sep. 2012. He was CEA Chair Professor from Oct. 2012 to Sep. 2017. He was the Head with the Department of Electrical Engineering, IIT Delhi from Jul. 2014 to Aug. 2016. He has been the Dean, Academics with the IIT Delhi from Aug. 2016 to Aug. 2019. He was JC Bose Fellow with DST, Government of India from Dec. 2015 to Jun. 2021. He is currently an Emeritus Professor and SERB National Science Chair at IIT Delhi since July 2021. He has guided 121 Ph.D. dissertations, and 182 M.E./M.Tech./M.S.(R) theses. He has 62 granted patents and 46 filed patents. He has executed more than ninety sponsored and consultancy projects. He has coauthored a textbook on power quality: *Power Quality Problems and Mitigation Techniques* (Wiley, New York, NY, USA, 2015). His research interests include solar PV grid interface systems, microgrids, power quality monitoring and mitigation, solar PV water pumping systems, improved power quality ac–dc converters, power electronics, electrical machines, drives, and electric vehciles.



Arunima Verma (Senior Member, IEEE) received the B.Tech. degree in electrical from Priyadarshini College of Engineering and Architecture, Nagpur University, Nagpur, India, in 2000, the M.Tech. degree in power systems from Malaviya National Institute of Technology Jaipur, Jaipur, India, in 2002, and the Ph.D. degree in electric drive and control from Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India, in 2011.

In 2003, she was an Assistant Professor with the Department of Electrical Engineering, Institute of Engineering and Technology, Lucknow. She became an Associate Professor in 2017 and a Professor in 2020. She has guided six Ph.D. dissertations and 12 M.E./M.Tech. thesis. Her areas of research interests include solar PV grid interface systems, microgrids, power quality mitigation, improved power quality ac–dc converters, and electric drives and control.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 04:51:24 UTC from IEEE Xplore. Restrictions apply.

Analysis of PCE-based Superplasticiser for the Different Types of Cement using Marsh Cone Test

Rahul Kumar[1], Manvendra Verma[1,*], Nirendra Dev[2]

[1]Department of Civil Engineering, GLA University, Mathura, Uttar Pradesh 281406, India

[2]Department of Civil Engineering, Delhi Technological University, Delhi, India

*Author to whom correspondence should be addressed:
E-mail: mv075415@gmail.com

(Received December 13, 2023; Revised March 22, 2024; Accepted April 17, 2024).

Abstract: The Marsh cone test is a direct and straightforward technique for examining the fluidity of cement paste using a superplasticiser. Polycarboxylate ether (PCE) based Auramix 400 superplasticiser, portland pozzolana cement, and ordinary Portland cement are used for finding the optimum dosage of superplasticiser with specific cement. The optimum dosage and the superplasticiser's efficiency are determined by the marsh cone test for two types of cement. The fresh and mechanical properties and non-destructive testing of the concrete mixes made by two cement types have been compared. The water-cement ratio is kept constant for both types of cement. Results show that the optimum dosage of superplasticiser depends upon the different types of cement. After the saturation point, the fluidity of cement paste doesn't affect increasing the quantity of superplasticiser. Both concrete mixes' fresh and mechanical properties of the same grade are similar, but the fine aggregate and coarse aggregate requirements are higher in OPC (Ordinary Portland Cement) mix concrete than in PPC (Portland Pozzolana Cement) mix concrete. The requirement of superplasticiser is more when used with PPC than the OPC concrete mix.

Keywords: Mechanical Properties; Marsh cone funnel; Workability; Superplasticizer; Saturation point.

1. Introduction

The need for good workability is the primary goal of the concrete. Different conditions necessitate varying degrees of workability in concrete. Structures where thin walls are required, beam-column joints, roof and foundation systems with congestion of reinforcement, deep beams, and columns or pillar junctions all need a high degree of workability [1,2]. In most circumstances, the simple solution is to use more water on the site, but excess water can harm concrete's strength and durability. Due to this issue, the SP (superplasticiser) should be used to enhance the workability of the concrete. The use of SP can permit water reduction by up to 30 percent without dropping the concrete's workability [3–5].

These are some properties of superplasticisers:

- It created significantly more workable concrete than plain concrete at the same w/c ratio.
- For the same workability, it allows the use of a lower w/c ratio.
- It decreases cement content due to greater strength with a reduced w/c ratio.

SP also creates homogeneous, cohesive concrete, i.e., free of segregation and bleeding.

The paste's flow behaviour and the dispersion of cement particles by the SP determine high-performance concrete's fresh properties. Various types and quantities of SP are employed to assess the flow properties of cement paste. “The viscous nature of the paste is characterized using the Bingham and Herschel-Bulkley models. The analysis reveals a consistent pattern across alterations in SP dosage, Marsh cone flow period, mini-slump spread, and rheological parameters” [6]. The marsh cone flow is investigated on glycerol-water mixtures using rheological measurements and digital image analysis [7–10]. The viscous Newtonian fluid used in the computations to solve the flow problem is discussed. It is discovered that flow time is relative to Newtonian viscosity, and the relationship between flow duration and apparent viscosity is only applicable for cement pastes with no yield stress [11]. The investigation is done on the effects of a polycarboxylate SP on the mechanical, mineralogical, microstructural, and rheological properties of Portland cement paste. The SP presence is found to slow down the initial cement hydration processes, and a low polycarboxylate SP dosage resulted in a significant reduction in yield stress [12]. The impact of water/cement ratio, mineral admixture composition, and SP type is analysed on cement pastes with varying SP dosages. It is discovered that there is an SP saturation dosage above which no substantial rise in

fluidity occurs [13]. Admixtures in concrete have certain potential benefits with the cement paste, such as air entrainment, water reduction, and plasticity [14]. Furthermore, the use of admixtures may be used to control the setting time, such as acceleration or retardation. Amounts of admixture added beyond a certain point can cause incompatibilities such as concrete's performance characteristics and setting time and strength gain issues [15–17]. The marsh cone analysis finds the admixture's optimum dosage for a given cement [18]. The study is conducted on how mineral admixtures and pozzolanic content are used to produce self-compacting concrete. After lowering the aggregate volume ratio, it is determined that fly ash and the proper quantity of SP are used to create the self-compacting concrete [19]. Polyethylene Glycol 4000 admixture is used to analyse self-compacting self-curing concrete and is found well-performed compared to conventional concrete [10]. "A Marsh cone coupled with a Plexiglas horizontal channel can produce a trend similar to that of a rheometer" [20]. The increase in water temperature resulted in a rise in flow time due to increased plastic viscosity and a faster hydration rate [21]. With increasing temperature, cement paste's yield stress and plastic viscosity improved. Since cement hydration reactions accelerate at higher temperatures, the rate of increase becomes steeper [22].

2. Materials

2.1 Cement

JK Laxmi Pro+ (fly ash) pozzolanic material-based PPC and Wonder cement OPC, both types of cement, are purchased from a material plant located alongside Delhi, India, which are used in this analysis[3,23]. The PPC and OPC are subjected to various methods of testing based on several IS codes (Indian Standard). The outcomes from the PPC and OPC are presented in Table 1.

Table 1. Characteristics of PPC (J.K Laxmi pro+ cement).

Cement type	Cement properties	Results
PPC	28 days compressive strength	42 MPa
	Normal consistency	33%
	Sp. Gr.	2.97
	Setting time (Initial)	90 min
	Setting time (Final)	155 min
	Cement expansion (soundness)	3 mm
OPC	28 days compressive strength	44 MPa
	Normal consistency	29.5%
	Sp. Gr.	3.14
	Setting time (Initial)	93 minutes
	Setting time (Final)	182 minutes
	Cement expansion (soundness)	2 mm

2.2 Aggregates

"FA (Fine aggregate) in the form of crushed stone (stone dust) and CA (Coarse aggregate) in 10 mm and 20 mm sizes are purchased from a building material plant located alongside Delhi, India. The testing of FA and CA are done within standard parameters specified in the IS code, and all outcomes associated with properties of aggregate are available in table 2. The sieve analysis method is used to find FA's gradation and particle size distribution, and the results are demonstrated in Fig 1" [14–17,24,25].

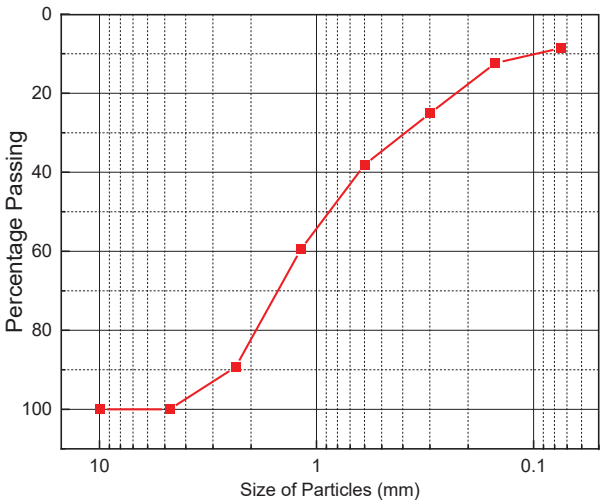


Fig. 1: Grain size distribution curve of fine aggregate

Table 2. Properties of FA and CA.

Material	Properties	Results
FA	Zone	II
	Fineness modulus	2.756
	Sp. Gr.	2.62
	Water absorption	1.21 %
CA	Sp. Gr.	2.79
	Water absorption	0.2 %
	Fineness modulus	7.29
	Impact value	26.3 %
	Loss angel abrasion Value	36 %

2.3 Superplasticiser

Fosroc Chemicals (India) Ltd. provides Auramix 400, a PCE (polycarboxylate ether) based SP that is used to provide concrete with the appropriate workability. SP has an Sp. Gr. of 1.09 and a pH value of 6, according to the Fosroc Chemicals. SP's optimal dose and efficiency are obtained with marsh cone equipment using a nozzle aperture diameter of 8 mm [26–28].

3. Mix Proportions

Based on trial results, a water-to-cement (w/c) ratio of 0.38 was chosen for the concrete in this investigation. The concrete's workability and compressive strength were

both evaluated as part of the examination. The Marsh cone test recommends a dose of 1.4 percent for Portland Pozzolana Cement (PPC) and 1 percent for Ordinary Portland Cement (OPC) in order to ensure a moderate degree of workability. The marsh cone test determines the optimum dosage of SP[4,5]. Proportions for the mix design of M40 grade concrete are presented in Table 3. Mix R1 is made by PPC, while R2 mix is made by OPC. In this analysis, a mechanical vibrator is used to achieve perfect compaction, minimise voids in the concrete, and ensure a consistent mix throughout. The concrete mixtures are prepared at a temperature of 29 degrees Celsius and under a relative humidity of 36%. Afterward 24 hours, the samples are demoulded and placed into the curing tank to achieve the strength of concrete at different ages [23,29]. The curing water is free from oils and other chemicals. The curing water is changed after every seven days to maintain the pH of the water. The mix proportion of M40 and all the materials used are presented in Table 3.

Table 3. Mix proportion for concrete.

Mix No.		**R1**	**R2**
Cement (kg/m^3)		390	390
FA (kg/m^3)		666	676
CA (kg/m^3)	(20 mm)	769.8	781.2
	(10 mm)	513.2	520.8
w/c ratio		0.38	0.38
SP (%)		1.4	1.0
Slump (mm)		97	99

4. Testing Procedure

Concrete's fresh and hardened properties are evaluated in this study, such as workability, compressive strength, and flexural strength non-destructive tests. For testing the hardened properties of concrete, three specimens are prepared for both sets of the mix. A UTM (universal testing machine) with a maximum loading capacity of 2000 tons is utilised to assess concrete's compressive and flexural strength. The specimen's dimension for compressive strength is 150X150X150 mm and for flexural strength is 100 mm x 100 mm x 500 mm as per IS code [30]. 140 kg/cm^2/min gradual loading rate is applied to the sample for testing compressive strength, and the 180 kg/min gradual loading rate is applied for flexural strength.

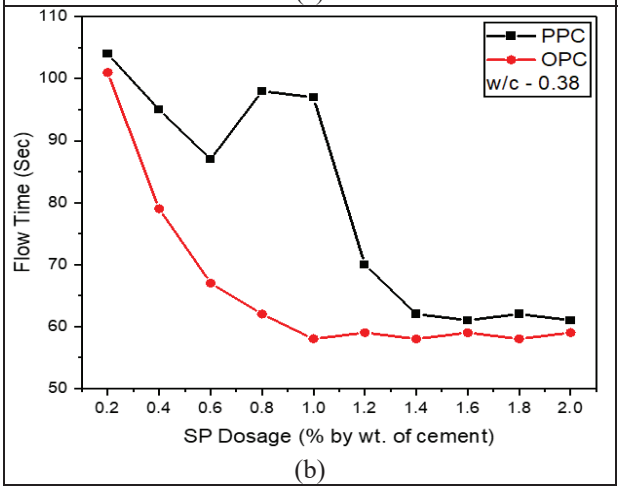
5. Results and Discussion

5.1 The Optimum dosage of superplasticizer

With the use of a marsh cone apparatus, the optimal amount and efficacy of SP are determined. The diameter of the nozzle aperture is 8 mm. There are many aperture sizes available, say 5, 6, 8, and 11 mm, to perform this test; however, the selection of the appropriate nozzle is based on cement paste fluidity. Other attachments with a bigger diameter aperture can be used according to different slurry types and different w/c ratios. In this test, w/c is taken 0.38 according to the prescribed mix design of concrete. The dosage of SP is taken 0.2% in the first test, then increased with an interval of 0.2% until the flow time becomes almost constant. Mix them thoroughly in the Hobart mixer for 2 min to avoid the lumps in the mixture. If hand mixing is done, then the slurry should pass through a 1.18 mm sieve before the marsh cone test to avoid lumps in the slurry. The SP's optimum dosage is selected when the slurry takes minimum time to complete out from the marsh cone. After the SP's optimum dosage (saturation point), the slurry's outcome time becomes almost constant for the particular cement and specified w/c ratio. After the saturation point, the addition of SP doesn't influence the fluidity of cement paste. The testing setup for determining the optimum dosage of SP is illustrated in Fig. 2(a). The slurry's flow time with different dosages is shown in Fig. 2(b). SP dosages of 1.4 percent for PPC and 1% for OPC are determined to be the most effective.



(a)



(b)

Fig. 2: a) Testing setup of marsh cone apparatus; **b)** The optimum dosage of superplasticizer with different cement

5.2 Efficiency of Superplasticiser

The marsh cone apparatus with an 8 mm aperture diameter is used for finding the SP's efficiency. The

efficiency of the SP is essential for making the mix design of desired strength. SP's efficiency indicates that the required amount of water for the concrete design mix can be reduced up to a specific limit. The test procedure for finding the SP's efficiency is taken from the book concrete technology (theory and practice)[31]. In this process, the cement quantity remains constant, but the water amount is changed in all the tests. The flow time of slurry without SP, having 1100 gm of water should be the same as another slurry with 1% of SP dosage having a less amount of water, as shown in Fig 3. The efficiency of the auramix-400 SP for PPC and OPC permits the decrement of 22.73% of water. Different kinds of SP may be used for the same nature of cement, and the SP efficiency will be dissimilar. To find the efficiency of the SP flow table test can be used in place of the marsh cone apparatus.

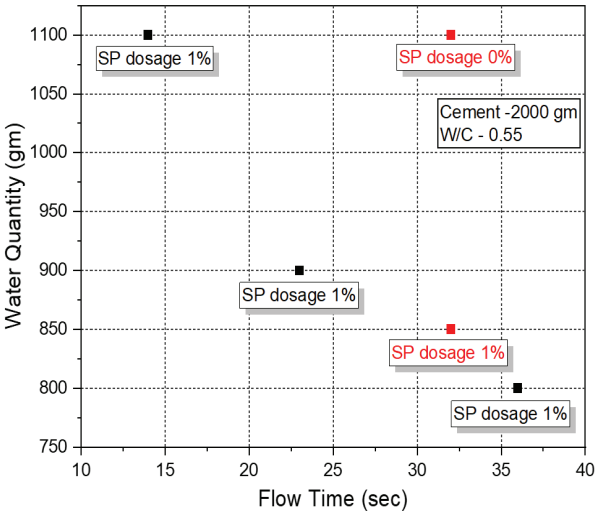


Fig. 3: The efficiency of SP

$$\text{The efficiency of SP} = \frac{\text{The difference in water content}}{\text{Original water content}} X100$$

$$= 22.73\%$$

5.3 Workability

The slump cone test measures the workability of concrete. SP is utilised by the weight of cement for ensuring the required workability of concrete. The slump cone results of the PPC and OPC mix are revealed in Fig 4. After the outcomes, it is assessed that the concrete workability of both mixes is nearly similar. The concrete mixes of the same grade made by PPC and OPC having the optimum dosage of SP exhibit similar workability. The slump value for medium workable concrete should be 50-100 mm, ensuring that both mixes are workable in the medium range.

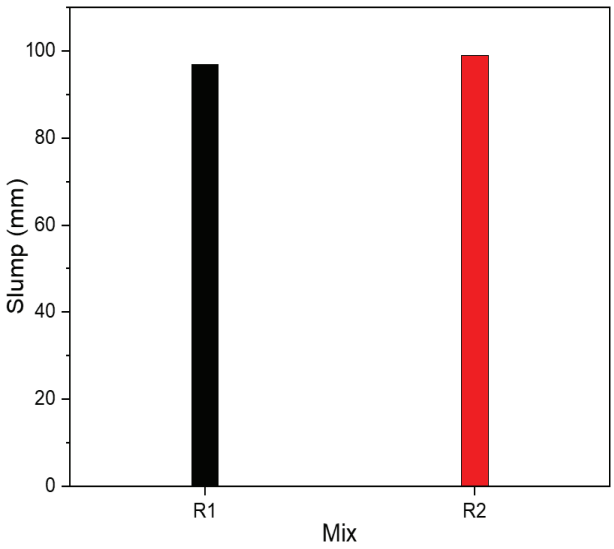


Fig. 4: Results of slump cone for mixes

5.4 Compressive Strength

The compressive strength of concrete is tested according to [32]. The compressive strength results of both mixtures are revealed in Fig 5. After analysing the results, it is found that the OPC mix's compressive strength is slightly larger at 28 days of curing and slightly lower at seven days and 14 days of curing as compared to the PPC mix. Both concrete mixes of the same grade show nearly similar results. The compressive strength of the PPC mix and OPC mix for the same SP would be almost similar, but the dosage of SP varies [33–38]. The dosage of SP in the PPC mix is slightly higher than the OPC mix, but the amount of FA and CA is lower.

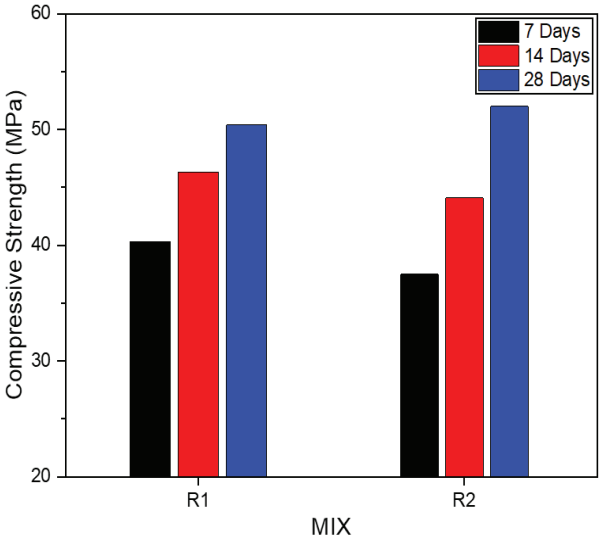


Fig. 5: Compressive strength of standard concrete of different mix

5.5 Flexural Strength

The Flexural strength of concrete is tested according to IS code [32]. The prism specimen's failure pattern is shown in Fig. 6(a), and the failure condition is in the range prescribed in IS code [32]. The results of the flexural strength of both mixtures are revealed in Fig. 6(b). Afterwards, evaluating the results, it is found that the OPC mix's flexural strength is slightly larger at 28 days of curing and slightly lower at seven days and 14 days of curing as compared to the PPC mix [39–41]. Both concrete mixes of the same grade show nearly similar results[42].

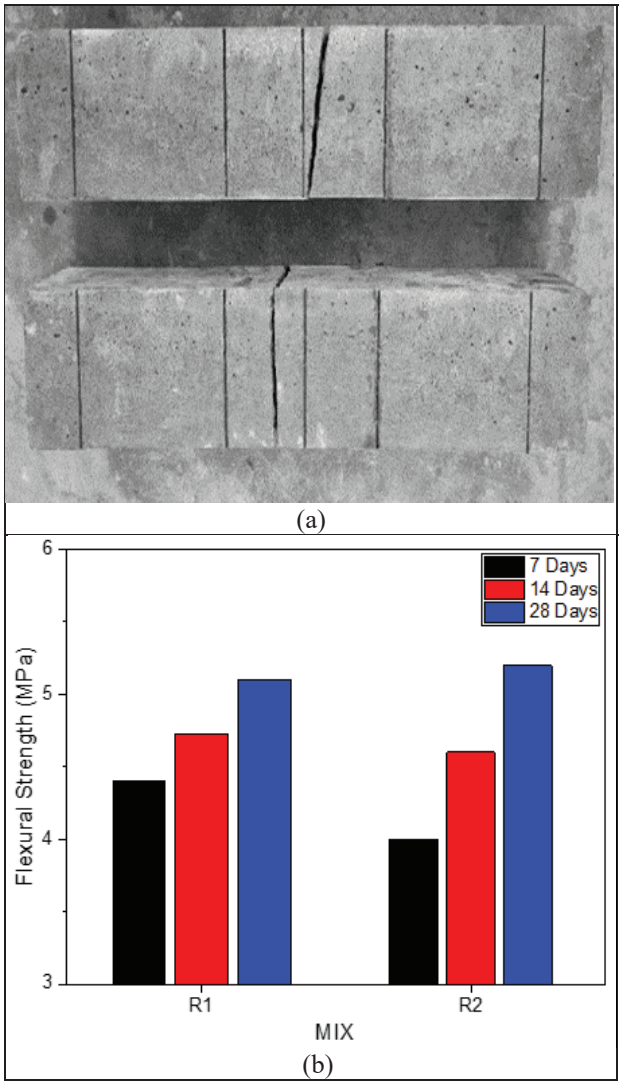


Fig. 6: a) Failure pattern of prism specimen; b) Flexural strength of standard concrete of different mix

5.6 Compressive strength of concrete mixes by Rebound hammer

Fig 7(a) shows the Proceq Schmidt Digital Concrete Test Hammer, utilised to assess the concrete's compressive strength according to [43]. The compressive strength results from the rebound hammer of both mixes are revealed in Fig.7(b). After examining the results, it is discovered that the compressive strength of concrete mixtures measured using a rebound hammer is 2-8 percent higher than the results obtained with UTM.

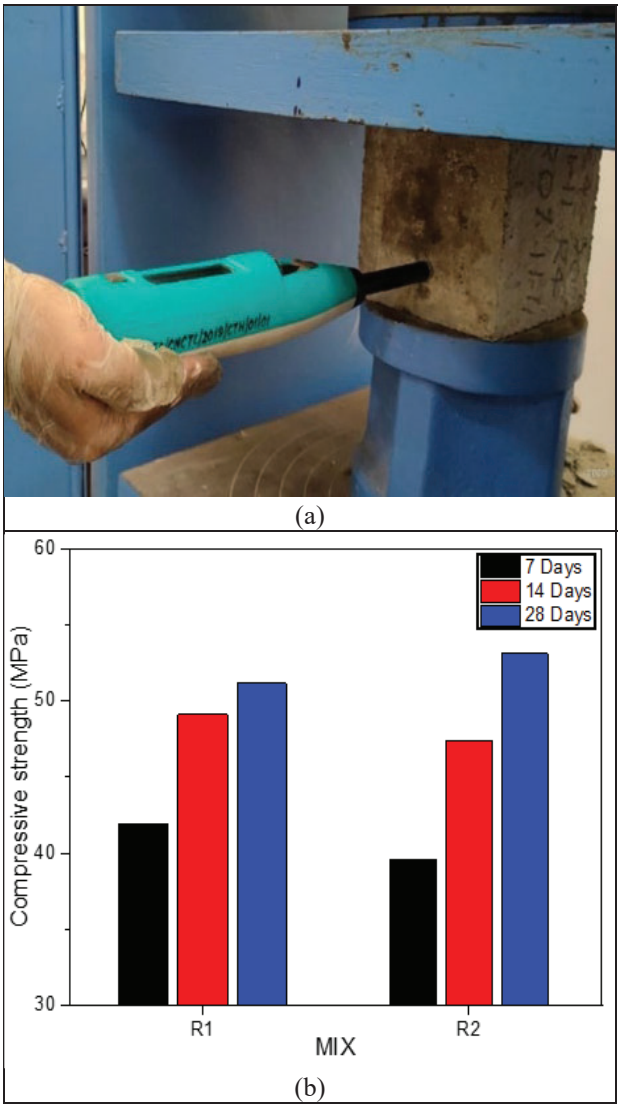


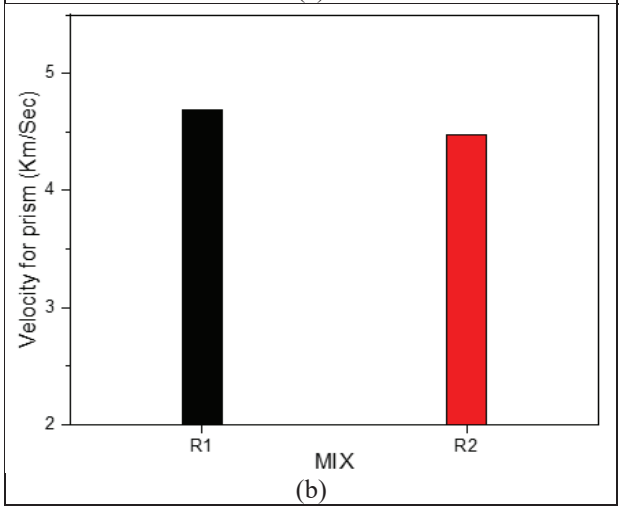
Fig. 7: a) Rebound hammer testing of the cube for the compressive strength test; **b)** Compressive strength of standard concrete by rebound hammer with age

5.7 Ultrasonic Pulse Velocity

The UPV of both concrete mixes is tested by a 'MATEST' UPV tester and 2 transducers with a frequency of 54 kHz on the prism specimens, as shown in Fig. 8(a) according to [44]. An adequate amount of gel is applied between the concrete prism's surface and the transducer to confirm the proper connection. The UPV results of both mixes are shown in Fig. 8(b). The PPC concrete mix has UPV greater than 4.5 km/sec, which shows that the concrete is in excellent condition, but the OPC concrete mix has UPV of less than 4.5km/sec, and it offers good condition [45,46].



(a)



(b)

Fig. 8: a) Ultrasonic pulse velocity test of standard concrete; **b)** Ultrasonic pulse velocity of mixes at 28 days

6. Conclusion

The optimum dosage of polycarboxylate ether-based SP auramix 400 is found 1.4% with PPC and 1% with OPC. The optimum dosage of SP depends upon the type of SP and type of cement. If the cement or SP is different, then the optimum dosage will be changed. The optimum dosage of SP also depends upon the water-cement ratio. If the water ratio is changed, the optimum dosage will be different. If the SP is used with PPC, it requires more SP than the requirement of SP with OPC. However, FA and CA's requirement is more in OPC mix concrete than PPC mix concrete. The fresh and mechanical properties of concrete mixes of the same grade of concrete made by PPC and OPC with the SP's optimum dosage are approximately similar. The UPV is slightly higher in PPC concrete than in OPC concrete.

Acknowledgement

We would like to express our gratitude to Delhi Technological University's Civil Engineering Department for their kind help in making this effort possible and improved. The calibre and extent of our efforts have been greatly enhanced by their priceless contributions.

Competing Interests
All authors have no financial interests.
Author contributions
The authors were involved in all parts of the manuscript preparation.

Nomenclature

FA- Fine aggregate
CA- Coarse aggregate
SP- Superplasticizer
PPC- Portland Pozzolana Cement
OPC- Ordinary Portland Cement
IS code- Indian Standard code
ASTM- American Society for Testing and Materials
Sp. Gr.- Specific gravity

References

1) G.G. Goswami, R. Sarwar, M. Rahman, D.C. Panday, I.J. Ishita, T. Labib, and N.V. Motiram, “Why did india pull out of regional comprehensive economic partnership (rcep)? a gravity explanation of the indian puzzle,” *Environ. Dev. Sustain.*, 10 (*03*) 1140–1155 (2023).
2) M. Sharma, N.L. Jain, and J.K. Purohit, “Analysis of circular economy barriers in manufacturing context for indian industries: a bwm ranking process,” *Environ. Dev. Sustain.*, 10 (*03*) 1156–1168 (2023). doi:10.1007/s10668-023-03868-9.
3) A. Chaudhary, “Road surface quality detection using light weight neural network for visually impaired pedestrian road surface quality detection using light weight neural,” *Evergr. Jt. J. Nov. Carbon Resour. Sci. Green Asia Strateg.*, 10 (*2*) 706–714 (2023).
4) B.R. Sathi, S.N. Gurugubelli, and H.B. N, “The effect of ecap on structural morphology and wear behaviour of 5083 al composite reinforced with red mud,” *Evergr. Jt. J. Nov. Carbon Resour. Sci. Green Asia Strateg.*, 10 (*02*) 774–781 (2023).
5) S. Nayak, and M.P. Kumar, “Mechanical characterization and static analysis of natural fiber based composite propeller blade mechanical characterization and static analysis of natural fiber based composite propeller blade,” *Evergr. Jt. J. Nov. Carbon Resour. Sci. Green Asia Strateg.*, 10 (*2*) 805–812 (2023).
6) C. Jayasree, and R. Gettu, “Experimental study of the flow behaviour of superplasticized cement paste,” *Mater. Struct. Constr.*, 41 (*9*) 1581–1593 (2008). doi:10.1617/s11527-008-9350-5.
7) M. Verma, K. Upreti, P. Dadhich, S. Ghosh, V. Khatri, P. Singh, and 1, “Prediction of Compressive Strength of Green Concrete by Artificial Neural Network,” in: ICACIS 2022, 2023: pp. 622–632. doi:/10.1007/978-3-031-25088-0_55.

8) M. Verma, "Prediction of compressive strength of geopolymer concrete by using ann and gpr," *Asian J. Civ. Eng.*, 24 (*8*) 2815–2823 (2023). doi:10.1007/s42107-023-00676-4.
9) U. Sharma, N. Gupta, and M. Verma, "Prediction of compressive strength of geopolymer concrete using artificial neural network," *Asian J. Civ. Eng.*, 24 (*8*) 2837–2850 (2023). doi:10.1007/s42107-023-00678-2.
10) M. Verma, and M. Nigam, "Mechanical behaviour of self compacting and self curing concrete," *Int. J. Innov. Res. Sci. Eng. Technol.*, 6 (*7*) 14361–366 (2017). doi:10.15680/IJIRSET.2017.0607245.
11) R. Le Roy, and N. Roussel, "The marsh cone as a viscometer: theoretical analysis and practical limits," *Mater. Struct.*, 38 (*1*) 25–30 (2005). doi:10.1007/bf02480571.
12) F. Puertas, H. Santos, and S. Palacios, Martinez-Ramrez, "Polycarboxylate superplasticiser admixtures: effect on hydration, microstructure and rheological behaviour in cement pastes," *Adv. Cem. Res.*, 17 (*2*) 77–89 (2005). doi:10.1680/adcr.17.2.77.65044.
13) L. Agulló, B. Toralles-Carbonari, R. Gettu, and A. Aguado, "Fluidity of cement pastes with mineral admixtures and superplasticizer - a study based on the marsh cone test," *Mater. Struct. Constr.*, 32 (*7*) 479–485 (1999). doi:10.1007/bf02481631.
14) R. Kumar, M. Verma, and N. Dev, "Investigation on the effect of seawater condition, sulphate attack, acid attack, freeze–thaw condition, and wetting–drying on the geopolymer concrete," *Iran. J. Sci. Technol. Trans. Civ. Eng. Civ. Eng.*, 46 (*4*) 2823–2853 (2022). doi:10.1007/s40996-021-00767-9.
15) R. Kumar, and N. Dev, "Effect of acids and freeze–thaw on the durability of modified rubberized concrete with optimum rubber crumb content," *J. Appl. Polym. Sci.*, 139 (*21*) 52191 (2022). doi:10.1002/app.52191.
16) R. Kumar, and N. Dev, "Assessment of mechanical and impact resistance properties of rubberized concrete after surface modification of rubber crumb," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, (2021). doi:10.1007/s40996-021-00784-8.
17) R. Kumar, and N. Dev, "Mechanical and microstructural properties of rubberized concrete after surface modification of waste tire rubber crumb," *Arab. J. Sci. Eng.*, 47 (*4*) 4571–4587 (2022). doi:10.1007/s13369-021-06154-w.
18) D. Jadhav, "Compatibility of chemical admixture with cement: marsh cone test," *Int. J. Adv. Mech. Civ. Eng.*, (*3*) 2394–2827 (2016).
19) I.Ö. Deneme, "Modelling of compressive strength of self-compacting concrete containing fly ash by gene expression programming," *Rev. La Constr.*, 19 (*2*) 346–358 (2020). doi:10.7764/RDLC.19.2.346.
20) M. Benaicha, O. Jalbaud, A. Hafidi Alaoui, and Y. Burtschell, "Marsh cone coupled to a plexiglas horizontal channel: rheological characterization of cement grout," *Flow Meas. Instrum.*, 45 126–134 (2015). doi:10.1016/j.flowmeasinst.2015.06.004.
21) M. Sonebi, M.T. Bassuoni, J. Kwasny, and A.K. Amanuddin, "Effect of nanosilica on rheology, fresh properties, and strength of cement-based grouts," *J. Mater. Civ. Eng.*, 27 (*4*) 04014145-1–11 (2015). doi:10.1061/(asce)mt.1943-5533.0001080.
22) D. Sathyan, A.K. Balakrishnan, and S.M. Mohandas, "Temperature influence on rheology of superplasticized pozzolana cement and modeling using rks algorithm," *J. Mater. Civ. Eng.*, 30 (*9*) 040182211–13 (2018). doi:10.1061/(asce)mt.1943-5533.0002406.
23) N.K. Yadav, N.S. Rajput, S. Kulshreshtha, and M.K. Gupta, "Investigation of the mechanical and wear properties of epoxy resin composite (ercs) made with nano particle tio2 and cotton fiber reinforcement," *Evergreen*, 10 (*1*) 63–77 (2023). doi:10.5109/6781041.
24) R. Kumar, M. Verma, N. Dev, and N. Lamba, "Influence of chloride and sulfate solution on the long-term durability of modified rubberized concrete," *J. Appl. Polym. Sci.*, 2022 (*139*) 1–15 (2022). doi:DOI: 10.1002/app.52880.
25) R. Kumar, N. Dev, S. Ram, and M. Verma, "Investigation of dry-wet cycles effect on the durability of modified rubberised concrete," *Forces Mech.*, 10 (*2023*) 100168 (2023). doi:10.1016/j.finmec.2023.100168.
26) M. Verma, and N. Dev, "Effect of Superplasticiser on Physical, Chemical and Mechanical Properties of the Geopolymer Concrete," in: Second ASCE India Conf. " Challenges Resilient Sustain. Infrastruct. Dev. Emerg. Econ., 2020: pp. 1183–1189.
27) M. Verma, and N. Dev, "Effect of liquid to binder ratio and curing temperature on the engineering properties of the geopolymer concrete," *Silicon*, 14 (*4*) 1743–1757 (2022). doi:10.1007/s12633-021-00985-w.
28) M. Verma, and N. Dev, "Effect of snf-based superplasticizer on physical, mechanical and thermal properties of the geopolymer concrete," *Silicon*, 14 (*3*) 965–975 (2022). doi:10.1007/s12633-020-00840-4.
29) M. Awi, and A.S. Abdullah, "A review on mechanical properties and response of fibre metal laminate under impact loading (experiment)," *Evergreen*, 10 (*1*) 111–129 (2023). doi:10.5109/6781057.
30) IS 10086 1982, "Specification for Moulds for Use in Tests of Cement and Concrete," 2008.
31) M.S. Shetty, "Concrete technology: theory and practice - google books," (2013).
32) IS: 1199 - 1959 Reaffirmed, "Methods of sampling and analysis of concrete," *Indian Stand.*, (2004).
33) A. Chouksey, M. Verma, N. Dev, I. Rahman, and K.

Upreti, "An investigation on the effect of curing conditions on the mechanical and microstructural properties of the geopolymer concrete," *Mater. Res. Express*, 9 (*5*) 55003 (2022). doi:10.1088/2053-1591/ac6be0.

34) M. Verma, N. Dev, I. Rahman, M. Nigam, M. Ahmed, and J. Mallick, "Geopolymer concrete: a material for sustainable development in indian construction industries," *Crystals*, 12 (*2022*) 514 (2022). doi:10.3390/cryst12040514.
35) M. Verma, K. Upreti, M.R. Khan, M.S. Alam, S. Ghosh, and P. Singh, "Prediction of Compressive Strength of Geopolymer Concrete by Using Random Forest Algorithm," in: ICACIS 2022, 2023: pp. 170–179. doi:10.1007/978-3-031-25088-0_14.
36) M. Nigam, and M. Verma, "Effect of nano-silica on the fresh and mechanical properties of conventional concrete," *Forces Mech.*, 10 (*22*) 100165 (2023). doi:10.1016/j.finmec.2022.100165.
37) M. Verma, K. Upreti, P. Vats, S. Singh, P. Singh, N. Dev, D.K. Mishra, and B. Tiwari, "Experimental analysis of geopolymer concrete : a sustainable and economic concrete using the cost estimation model," *Adv. Mater. Sci. Eng.*, 2022 1–16 (2022). doi:10.1155/2022/7488254.
38) K. Upreti, M. Verma, M. Agrawal, J. Garg, R. Kaushik, C. Agrawal, D. Singh, and R. Narayanasamy, "Prediction of mechanical strength by using an artificial neural network and random forest algorithm," *J. Nanomater.*, 2022 1–12 (2022). doi:DOI: 10.1155/2022/7791582.
39) A. Shahid Multani, and P.K. Gupta, "Investigation of pull-out characteristics of connections for post-installed rebar utilizing mortar-based binders and chemical adhesives," *Aust. J. Struct. Eng.*, (2024). doi:10.1080/13287982.2023.2247288.
40) P.K. Gupta, R. Kumar, Y.K. Gupta, and P.K. Mehta, "Effect of acidic environment on self compacting concrete," *Int. J. Civ. Eng. Technol.*, 8 (*2*) 595–606 (2017).
41) S. Saxena, L.B. Roy, P.K. Gupta, V. Kumar, and P. Paramasivam, "Model tests on ordinary and geosynthetic encased stone columns with recycled aggregates as filler material," *Int. J. Geo-Engineering*, 15 (*1*) 1–13 (2024). doi:10.1186/s40703-023-00202-0.
42) M. Sharma, N.L. Jain, and J.K. Purohit, "Analysis of circular economy enablers in manufacturing context for indian industries: a electre method ranking process," *Evergreen*, 10 (*3*) 1156–1168 (2023). doi:10.5109/7148437.
43) Bureau of Indian Standards, "Method of non-destructive testing of concrete, is 13311-(part 2) 1992, reaffirmed 2004, rebound hammer," *Bur. Indian Stand. New Delhi*, (2004).
44) IS 13311 Part 1 - 1992 Reaffirmed, "Method of non-destructive testing of concrete, part 1: ultrasonic pulse velocity," *Bur. Indian Stand.*, 1–7 (2004).
45) M. Verma, and N. Dev, "Effect of ground granulated blast furnace slag and fly ash ratio and the curing conditions on the mechanical properties of geopolymer concrete," *Struct. Concr.*, 23 (*4*) 2015–2029 (2022). doi:10.1002/suco.202000536.
46) M. Verma, and N. Dev, "Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete," *Struct. Concr.*, 22 (*S1*) E368–E379 (2021). doi:10.1002/suco.202000068.

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10574984

Analyzing Yoga Pose Recognition: A Comparison of MediaPipe and YOLO Keypoint Detection with Ensemble Techniques

Debashree Debalaxmi
Department of Information Technology,
Delhi Technological University
Bawana Road, Delhi-110042, India
debashreedebalaxmi@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology,
Delhi Technological University
Bawana Road, Delhi-110042, India
dvishwakarma @gmail.com

Virender Ranga
Department of Information Technology,
Delhi Technological University
Bawana Road, Delhi-110042, India
drvirender.ranga@gmail.com

***Abstract*— The practice of yoga encompasses mental, physical, and spiritual dimensions, aiming for holistic well-being. Accurate alignment in yoga enhances the effectiveness of each pose by targeting specific muscle groups, reducing strain on muscles and joints, and improving stability and balance. This research employs advanced computer vision techniques, YOLO (You Only Look Once) and MediaPipe to identify critical keypoints from the skeletal structures of yoga practitioners, thereby providing a detailed representation of body alignment for posture recognition. Augmented using the SMOTE technique, the skeletal data serves as input for various Machine Learning and ensemble models during the training process. The study utilizes a 2D image dataset comprising 20 well-known yoga poses. Among the models tested, the LightGBM ensemble classifier using MediaPipe keypoints achieved the highest accuracy at 96.52%. Further analysis included the evaluation of the model through a confusion matrix, learning curve, and pose-wise accuracy, even for similar-looking exercises. These findings highlight the potential of integrating computer vision and machine learning to enhance yoga practice through precise posture recognition and alignment analysis.**

Keywords—pose recognition, yoga, machine learning, MediaPipe, YOLOv8

I. INTRODUCTION

In today's world, the pursuit of holistic well-being has grown in importance, and yoga serves as a beacon, offering a means to find balance amidst the chaos. By adhering to correct postures, techniques, and sequences, yoga practitioners can maximize health benefits. Various methods to learn yoga include attending classes at yoga centers, watching videos, looking at images, or reading books. Given the fast-paced lifestyles of many, practicing yoga at home has become a preferred option. However, it can be challenging for individuals to identify flaws in their own postures. Incorrect postures can lead to various health issues such as joint pain, muscle strain, and sprains. Therefore, ensuring proper technique is crucial to avoid these potential problems and to fully reap the benefits of yoga [1]. Therefore, to assist people to perform yoga in an effective manner, there is a need for scientific evaluation of improper yoga posture identification and rectification by providing feedback. To support self-learning, this research tackles a critical component of self-learning in yoga practice, which is the requirement for precise yoga position recognition.

This research study has incorporated keypoint detection to construct a skeletal framework, thereby advancing the accuracy and depth of yoga pose detection. This study has conducted a comparative analysis between two leading methodologies, YOLOv8 and MediaPipe [2], to assess their accuracies in the context of yoga pose detection. YOLOv8 is well-known for its real-time object identification, which performs accurately, and MediaPipe, a precision-focused framework, performs exceptionally well. To mitigate imbalanced classes, SMOTE oversamples minority classes to provide a representation of postures. Next, the yoga positions is classified using a variety of machine learning techniques, namely Naïve Bayes, Support Vector Machine, KNN and Decision Trees. Afterwards, different ensembling classifiers like LightGBM, Random Forest, XGBoost and Bagging are used to improve the model performance. Various accuracy levels offer useful comparison points for assessing how well various approaches work in the particular area of yoga position recognition. This programme helps people practice yoga without the need for a trainer. The proposed system's dependability and performance are enhanced through this iterative process, ensuring that practitioners may rely on its guidance for the safest and most optimal yoga experience.

II. RELATED WORKS

To gain knowledge on various techniques for estimating yoga poses, some of the most recent works in this field are studied. A variety of techniques have been employed to estimate yoga posture. Skeletal and non-skeletal approaches are commonly employed in yoga pose recognition and classification systems. Skeletal yoga pose recognition identifies joint positions using models like OpenPose, PoseNet etc while non-skeletal classification recognizes poses based on overall visual features using architectures like ResNet, DenseNet etc.

Chen et al. [3] proposed a yoga training system in their research that studies participant's stances using the Y-system. Prominent axes, skeleton-based characteristics, and contour-based interest points were all incorporated in this system. For performance comparison, Microsoft Kinect and the OpenNI library were used to build skeletal diagram for yoga positions. The system used a depth sensor and RGB camera to gather depth data for tracking and estimating human body posture. This approach was good, but its fundamental drawback was the exorbitant cost and lack of user-friendliness nature of the Kinect device.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:42:56 UTC from IEEE Xplore. Restrictions apply.

Extreme Gradient Boosting (XGBoost) was applied to the OpenPose tool by Kakulapati et al. [4] in order to categorize photos of five different stances. They divided the dataset where 80% were reserved to train the model, while the remaining is for testing, treating the X and Y coordinates as crucial features. With an accuracy of 88%, the XGBoost classification earned the greatest result.

Ajay et al. [5] included five yoga postures, namely Natarajasana, Virbhadrasana 1, Virbhadrasana 2, Vrikshasana, Trikonasana, and Utkatasana in a dataset that was assembled using web resources and contributions from system developers. In order to identify posture mistakes, the proposed deep learning approach incorporated a human joints localization model into its convolutional neural networks (CNN) for yoga posture identification. The limitation of this research was the dataset used which comprises of a very few categories.

In order to create a random forest classifier for evaluating yoga asanas, Mustafa et al. [6] used transfer human learning position estimation techniques to identify 136 body-wide essential points. They use the popular posture estimation method AlphaPose in the first stage and then a random forest classifier in the second. They created a brand-new dataset that emphasizes different viewpoints in order to categorize yoga poses. They relied on accurate keypoint recognition and posture estimation.

In a work published in 2022, Ujjayanta et al. [7] presented a trained model built using the COCO keypoint identification dataset technique. A well-known search engine was used to automatically generate the dataset, which was then divided keeping three-fourth of it in train sets. However, a more extensive dataset is required.

Debabrata et al. [8] proposed a system for tracking the different yoga stances. They extracted the 33 keypoints obtained from MediaPipe package and followed by training of a combined classifier of CNN and LSTM. Vivek et al. [9], developed a deep learning-based Yoga position estimation system that can recognize posture while also providing comments on how to improve it. A total of 18 keypoints are extracted from the dataset. Nevertheless, imprecise body map segmentations lead to errors in displayed guidelines.

Ranjana et al. [10] provide a system for recognizing and correcting yoga postures that makes use of the OpenCV computer vision module as well as a kinematic description of the body. On a 7-yoga position dataset, the system uses 17 keypoints mapped using a method known as tf-pose estimation. This kypoint recognition model was developed with Keras and then trained with TensorFlow's MoveNet.

Janardhana et al. [11] used deep learning techniques to recognize yoga positions, allowing practitioners to monitor and improve their form. A combination of OpenCV and MediaPipe architecture has been used to detect important body parts, allowing yoga poses to be classified and feedback to be provided in real time. Their proposed model exceeds performance of cutting-edge techniques with 99.53% accuracy using a dataset of six yoga poses.

Deepak et al. [12] provide a deep learning approach based on MediaPipe and MoveNet for automatic yoga stance recognition. Our method extracts each person's key points and skeletons from input yoga photos by utilizing MediaPipe for key point-based skeletonization. After processing, the photos are sent into the MoveNet model, which guarantees accurate identification of the yoga stance.

To identify five different yoga positions with minimal data, Ashraf et al. [13] suggested a deep learning method inspired from Xception model. Their YoNet architecture deliberately leverages dimensional and spatial information out of the picture and uses them to aid with classification.

The previous research generally concentrated on a very limited set of well-liked yoga poses, often taking into account about five to seven popular poses. Moreover, those chosen poses differ significantly from one another in terms of body alignment, resulting in higher accuracy for the recognition models. However there are numerous variations of yoga which share similarities in appearance for which we can actually test the model's robustness and generalizability. More diversified datasets are thus required in order to train and evaluate the systems. Also, due to the inherent variety of yoga poses and similarity in movements, yoga pose classification may benefit from ensemble learning approaches, which has seen limited use in this area.

III. Proposed Approach

To properly categorize yoga postures, the suggested method includes dataset preprocessing, skeletal keypoint extraction, data augmentation, training of machine learning as well as ensemble models and then followed by testing and model evaluation as shown in

Figure **1**.

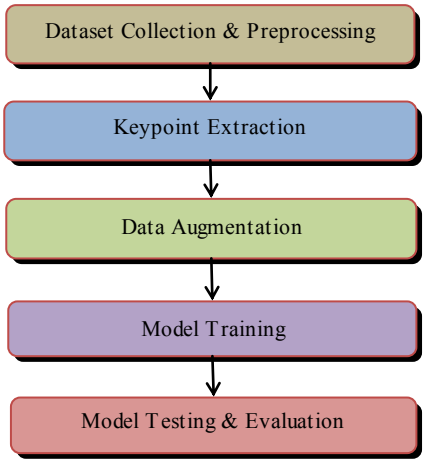


Figure 1 Proposed Approach

A. Dataset Collection and Preprocessing

We carefully selected a total of 1248 images of 20 yoga postures, each with a range of 50 to 90 photos, using the dataset we obtained from Kaggle [14]. We present a consolidated dataset with a variety of yoga poses from varied backdrops, lighting conditions, and viewpoints. Important meta data is attached to every image, including the file path, pixel values, and label that corresponds to the asana index. The yoga poses used here are 'Virabhadrasana i', 'Vajrasana', 'Savasana', 'Balasana', 'Bakasana', 'Setu bandha sarvangasana', 'Salabhasana', 'Vriksasana', 'Paschimottanasana', 'Uttanasana', 'Trikonasana', 'Virasana', 'Tadasana', 'Matsyasana', 'Virabhadrasana ii', 'Ardha matsyendrasana', 'Kapotasana', 'Adho mukha svanasana', 'Bhujangasana' and 'Dhanurasana'. The proposed method

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:42:56 UTC from IEEE Xplore. Restrictions apply.

seeks to include a greater variety of poses than existing work, including poses with similar alignments, like 'vajrasana' and 'virasana'. This varied dataset will aid in better understanding for evaluation of the models as they may misclassify and get confused.

To guarantee consistency and quality, certain preprocessing measures are implemented as soon as the dataset is gathered. These include image scaling, normalization, and label encoding. A split of 80:20 is followed, which means 80% data is used to train the model, and 20% is reserved for testing the model.

B. Keypoint Extraction

After preprocessing, two distinct keypoint extraction methods, namely YOLOv8 and MediaPipe are applied to extract skeletal keypoints from the images [14] as depicted in

Figure 2. These landmarks stand in for important anatomical regions, including the shoulders, wrists, elbows, knees, hips, and ankles. These landmarks' x, y, and z coordinates are also kept in arrays.

YOLOv8 functions through developing a grid based upon the input image. It enables fast and precise object detection via the use of only one neural network to forecast several bounding boxes and the class probabilities that correspond to them for every cell in the grid.

MediaPipe utilises deep learning algorithms trained on massive datasets to locate and monitor keypoints of interest, such as facial marking or joints on the human body. Convolutional neural network is used to identify traits from the input images and estimate the coordinates of landmarks, enabling continuous monitoring and evaluation.

These two keypoint extraction methods have been used due to their popularity in multiple classification tasks [2], [15].



Figure 2 Landmark detection in a yoga image [14]

C. Data Augmentation

A variety of methods, including rotation, scaling, flipping, cropping, translating, adding noise, varying brightness and contrast, color jittering, random erasing, and cutting, are included in data augmentation. By producing synthetic examples, these techniques are used to diversity the dataset, improve the model's capacity for generalization, and lower the likelihood of overfitting. It also has the benefit of permitting the model to perform generalisation on future unidentified images.

One method that is widely used for this is SMOTE, which is an abbreviation for 'Synthetic Minority Over-sampling Technique'. By interpolating between the current instances and their closest neighbors, SMOTE creates artificial training records for the minority class [16]. For each observation, after determining the k-nearest neighbors, we generate synthetic data points by interpolating between them. First, a stochastic sample is used to select the minority class. Then, SMOTE oversamples the minority class, and due to this successfully addresses class imbalance, decreases overfitting, and enhances model performance.

D. Model Training

The skeletal data is used to train a variety of machine learning and ensemble algorithms, to identify underlying patterns between the extracted keypoints and yoga position labels [17]. GridSearchCV is employed to methodically adjust the models' hyperparameters to find the optimal values which provide the best results during prediction. It operates by sifting through a predetermined set of parameters and assessing the effectiveness of each combination using cross-validation. The models that were utilized are as follows.

Linear SVM: The linear support vector machine (SVM) [18] determines the ideal hyperplane that maximizes margin between various stance classes. High-dimensional data, such as keypoints retrieved from pose photos, benefit greatly from its use.

Decision Tree: Decision Trees are trained to recursively partition the feature space based on landmark values in the classification of yoga poses .

K-Nearest Neighbors (KNN): KNN uses a distance measure between feature vectors for each sample [19]. This makes precise categorization based on important body positions and angles.

Multi-Layer Perceptron (MLP): MLP uses hidden layers in their multi-layered network to recognize complex details. In our context, the complexity comes from the variety of yoga postures in the dataset.

Gaussian Naive Bayes: It determines the probability that a given pose falls into a given category using the given keypoints. It simplifies the estimate of mean and standard deviation per pose by assuming a Gaussian data distribution.

Random Forest (RF): Random Forest builds and applies different feature subsets to train various decision trees. Every tree makes an individual pose class prediction during classification, and the final classification is the result of combining these predictions .

Bagging: Bagging improves accuracy by bootstrapping data. This means selecting subsets of data sampled with replacement. Each subset is also simultaneously trained on classifiers. Predictions are then aggregated to produce insights that increase robustness and accuracy.

eXtreme Gradient Boosting (XGBoost): It follows a boosting strategy of successively creating an ensemble of decision trees. It also handles missing data and is a well-used option because of its advantages which include maximizing precision and speed [20].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:42:56 UTC from IEEE Xplore. Restrictions apply.

LightGBM: LightGBM builds an ensemble of decision trees sequentially. It uses histogram-based techniques and leaf-wise tree growth, unlike traditional decision trees based gradient boosting. This results in a faster and more precise model while also lowering memory usage.

E. Model Testing & Evaluation

After rigorous training phase, model performance is evaluated using the test set. Various measures like test accuracy, F1-score, precision, recall and ROC-AUC score are determined. To gain more insight into the performance of each model in relation to the landmark detection methods employed, a comparative study is performed. The weights of the model that outperformed the others in terms of accuracy highest test accuracy are then saved.

IV. Experimental Analysis

A. Requirements and Specification

The experimenting method was conducted using Python 3.10 to guarantee dependability. Various machine learning and ensemble models were trained on the provided dataset, which comprises 20 classes representing different yoga poses. The hardware prerequisites for these models include Tesla T4 GPU, two virtual CPUs, 32GB of RAM, all accessible via Google Colab's network.

B. Experimental Study and Results

For pose estimation, YOLOv8 detects 32 keypoints, while the MediaPipe framework detects 33 keypoints. After training on each set of skeletal data, a list of optimal hyperparameters for each model is provided in Table I. These were obtained using Grid Search technique and the given values yielded the best performance.

TABLE I OPTIMAL HYPERPARAMETERS DETECTED

Model	Parameter	YOLOv8	MediaPipe
Linear SVM	penalty	L1	L2
	loss	hinge	hinge
	multi_class	ovr	ovr
Decision Tree	min_samples_leaf	2	1
	max_depth	40	30
	criterion	entropy	entropy
	splitter	best	best
K-Nearest Neighbors	p	2	2
	n_neighbors	3	3
	algorithm	auto	auto
Multi-Layer Perceptron	activation	relu	relu
	solver	adam	adam
	early_stopping	True	False
GaussianNB	priors	None	None
	var_smoothing	1e-09	1e-09
Random Forest	bootstrap	False	False
	max_features	sqrt	sqrt
	criterion	gini	entropy
	oob_score	False	False
Bagging	oob_score	False	False
	n_estimators	164	100
XGBoost	max_depth	8	6
	learning_rate	0.005	0.001
LightGBM	learning_rate	0.005	0.001
	boosting_type	gbdt	rf
	num_leaves	31	31

The accuracy of trained models for yoga position categorization utilizing keypoints found by MediaPipe and YOLOv8 is summarized in the Table II. All things considered, MediaPipe routinely beats YOLOv8 for all the models trained. With YOLOv8 keypoints, the LightGBM obtains the best accuracy (82%) but with MediaPipe landmarks, it shows remarkably higher accuracy (96.52%), closely followed by Random Forest (95.65%). All other ensemble methods also showed consistently good accuracy of over 90% using MediaPipe keypoints. Multi-Layer Perceptron also exhibits notable accuracy gains with both YOLOv8 and MediaPipe keypoints. It is also observed that ensemble learning methods consistently outperformed the base ML models.

TABLE II ACCURACY TABLE WITH RESPECT TO TWO DIFFERENT KEYPOINT DETECTION APPROACHES

Base Models	Accuracy (YOLOv8)	Accuracy (MediaPipe)
Linear SVM	65.28%	85.65%
Decision Tree	64.94%	81.73%
K-Nearest Neighbors	70.00%	70.00%
Multi-Layer Perceptron	80.00%	87.39%
GaussianNB	65.45%	88.69%
Random Forest	81.00%	95.65%
Bagging	80.94%	94.78%
XGBoost	79.28%	93.91%
LightGBM	**82.00%**	**96.52%**

Table III shows performance metrics like precision, recall, F1-score of all the models trained with MediaPipe keypoints. With LightGBM being the best, other ensembles-Random Forest, XGBoost and Bagging also exhibit remarkable results, with precision, recall and F1-scores over 0.94. Looking at Table II and TABLE III, overall it was discovered that LightGBM with MediaPipe keypoints was the best model.

TABLE III PERFORMANCE METRICS WITH RESPECT TO MEDIAPIPE

Algorithm	Precision	Recall	F1 score
Linear SVM	0.86	0.85	0.85
Decision Tree	0.82	0.81	0.81
K-Nearest Neighbors	0.70	0.69	0.69
Multi-Layer Perceptron	0.87	0.87	0.86
GaussianNB	0.90	0.89	0.88
Random Forest	0.96	0.96	0.96
Bagging	0.95	0.95	0.95
XGBoost	0.94	0.94	0.94
LightGBM	0.97	0.96	0.96

Figure 3 shows the confusion matrix, wherein a visualization of pose-wise accuracy of each yoga pose is shown. The diagonals in the matrix counts the number of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:42:56 UTC from IEEE Xplore. Restrictions apply.

correct predictions. 14 poses, namely 'Adho mukha svanasana', 'Ardha matsyendrasana', 'Bhujangasana', 'Kapotasana', 'Paschimottanasana', 'Salabhasana', 'Savasana', 'Tadasana', 'Uttanasana', 'Vajrasana', 'Virabhadrasana i', 'Virabhadrasana ii', 'Virasana', 'Vriksasana' show an accuracy of 100%, meaning all test images were classified correctly. The other elements in the matrix denote incorrect classifications for every exercise. It is because some poses are too intricate to be photographed from one single shot. Some poses carry many similarities like how the torso is oriented and the legs are positioned. For instance, in the 9th row, 'matsyasana' has been correctly predicted 7 times and misclassified as 'bhujangasana' once. This visualization helps in identifying problematic classes and defining scope for future improvements. Accuracy of the other yoga poses are: 'bakasana' (93.33%), ;setu bandha sarvangasana' (93.33%), 'matsyasana' (87.5%), 'trikonasana' (87.5%), 'balasana' (85.71%) and 'dhanurasana' (77.77%). For the exercises where comparatively lower accuracy is obtained, more images can be added to the dataset taken from various angles and direction for overall understanding of the pose by the model.

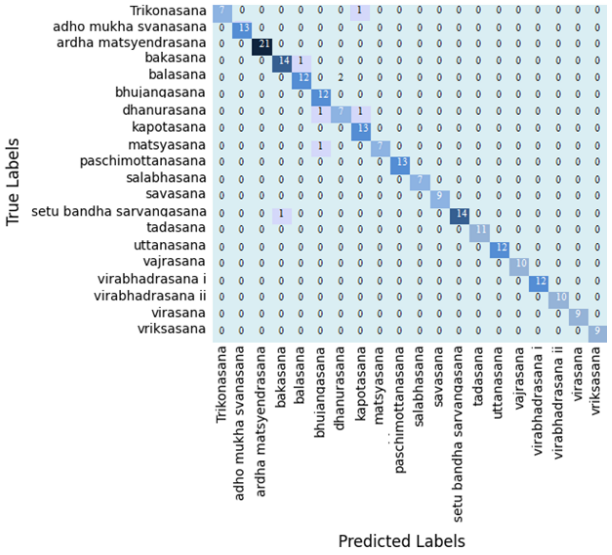


Figure 3 Confusion Matrix using MediaPipe+LightGBM model

In Figure 4 shows the accuracy and loss graphs of the MediaPipe+LightGBM model. The number of epochs, or one whole run across the training set, is shown on the x-axis. The left graph's y-axis denotes accuracy whereas the right one's represents loss. The accuracy curve starts at around 0.4 and increases to nearly 0.96 over the 25 epochs. This shows that the predictive algorithm is boosting its rate of accuracy in identifying the data points. The initial high loss (around 3.0) signifies significant discrepancies across the predictions to the true data points. The loss steadily decreasing towards 0 over 25 epochs indicates the model is progressively refining its predictions to better match the actual data.

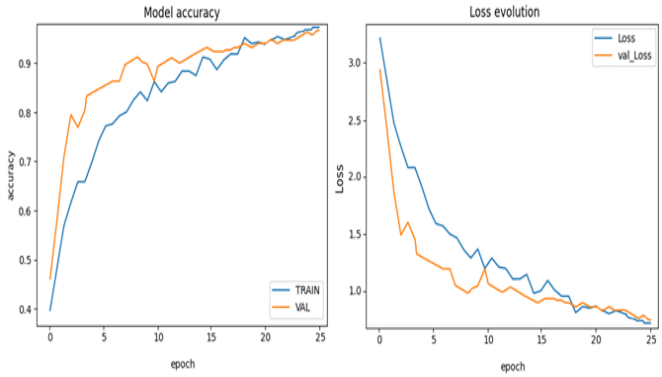


Figure 4 Accuracy and Loss curves using MediaPipe+LightGBM model

The learning curve of the LightGBM model with MediaPipe keypoints is shown in Figure 5, where the x-axis represents the count of training images and y-axis represents the cross-validation score, which rises quickly from roughly 0.65 to 0.96 as more training samples are added. Beyond a score of 0.96 improvement flattens and settles around a remarkable score of 0.97, which signifies a well-trained model. Also, the flattening of the curve indicates that after around 1200 samples, the model performance stays the same. As more samples are used for training, the shaded region encompassing the cross-validation score line diminishes, signifying a drop in uncertainty in the model's performance.

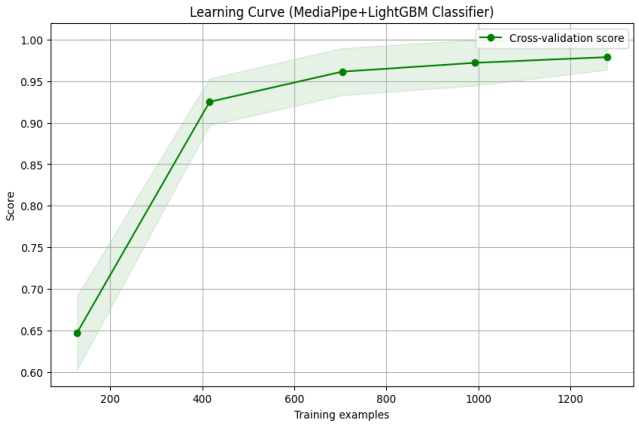


Figure 5 Learning Curve for LightGBM model

A graph of Receiver Operating Characteristic (ROC) is also shown for each class in Figure 6. AUC, which is the Area Under the Curve, quantifies classifier performance, with 1.0 indicating perfection classification. Classes like 'Trikonasana', 'adho mukha svanasana', and 'ardha matsyendrasana' have AUCs of 1.0. 'Bakasana' and 'setu bandha sarvangasana' achieve AUCs of 0.99, indicating high performance. Analysis of ROC curves and AUC values guides identification of classes necessitating improvement or more training data.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:42:56 UTC from IEEE Xplore. Restrictions apply.

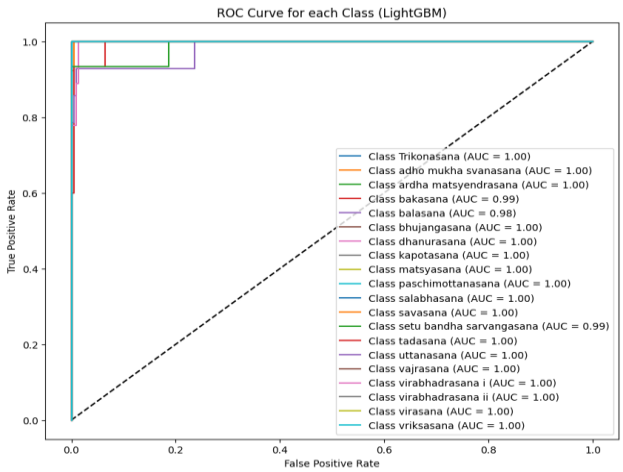


Figure 6 AUC-ROC curve using MediaPipe+LightGBM model

C. *Class prediction*

The true and predicted labels with landmarks indicated for pictures from the testing dataset showing different yoga positions are shown in Figure 7. From the six test images, five of them are correctly classified and one image of 'Trikonasona' pose is misclassified as 'kapotasana'.



Figure 7 True vs Predicted labels of some test images of poses [14]

V. Conclusion and Future Work

In summary, this study focuses on using YOLOv8 and MediaPipe keypoint detection for classifying yoga poses. Classification was performed using multiple base models and ensemble learning models on a custom dataset of varied yoga poses. For accurate posture identification, models using MediaPipe keypoints regularly outperformed their YOLOv8 counterparts. These results show how crucial accurate keypoint identification is to enhancing classification performance. Further, ensemble learning classifiers consistently showed better performance compared to base models which show the power of ensemble algorithms in collectively improving performance. LightGBM using MediaPipe keypoints showed the best performance. This model was further analyzed using to confirm the authenticity and correctness of results. Moreover, with this approach we can see that similar looking poses are also being distinguished and classified almost correctly.

In the future, more complex deep learning architectures such as CNNs or GNNs can be used for pose categorization. Also, using real-time inference, the proposed system can be enhanced to get interactive feedback. Also, more number of yoga poses should be considered to make a standard and large dataset. Moreover, the yoga posture classification task can be made more accessible when user-friendly mobile applications are developed.

References

[1] F. Rishan, B. De Silva, S. Alawathugoda, S. Nijabdeen, L. Rupasinghe, and C. Liyanapathirana, 'Infinity Yoga Tutor: Yoga Posture Detection and Correction System', in *2020 5th International Conference on Information Technology Research (ICITR)*, Dec. 2020, pp. 1–6. doi: 10.1109/ICITR51448.2020.9310832.

[2] H. Boudlal, M. Serrhini, and A. Tahiri, 'A novel approach for simultaneous human activity recognition and pose estimation via skeleton-based leveraging WiFi CSI with YOLOv8 and mediapipe frameworks', *Signal Image Video Process.*, vol. 18, no. 4, pp. 3673–3689, Jun. 2024, doi: 10.1007/s11760-024-03031-5.

[3] H.-T. Chen, Y.-Z. He, C.-C. Hsu, C.-L. Chou, S.-Y. Lee, and B.-S. P. Lin, 'Yoga Posture Recognition for Self-training', in *MultiMedia Modeling*, vol. 8325, C. Gurrin, F. Hopfgartner, W. Hurst, H. Johansen, H. Lee, and N. O'Connor, Eds., in Lecture Notes in Computer Science, vol. 8325. , Cham: Springer International Publishing, 2014, pp. 496–505. doi: 10.1007/978-3-319-04114-8_42.

[4] V. Kakulapati, P. sai Nalluri, and S. Kokkirala, 'Xgboost Analyses of Classification of Yoga'. Rochester, NY, Oct. 22, 2021. doi: 10.2139/ssrn.3948127.

[5] A. Gupta and A. Jangid, 'Yoga Pose Detection and Validation', in *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA)*, Goa, India: IEEE, Sep. 2021, pp. 319–324. doi: 10.1109/IRIA53009.2021.9588714.

[6] M. Chasmai, N. Das, A. Bhardwaj, and R. Garg, 'A View Independent Classification Framework for Yoga Postures', *SN Comput. Sci.*, vol. 3, no. 6, p. 476, 2022, doi: 10.1007/s42979-022-01376-7.

[7] U. Bhaumik, K. K. Singh, A. S. Akbari, and M. K. Bajpai, 'A Deep Learning-Based Approach to Detect Correct Suryanamaskara Pose', *SN Comput. Sci.*, vol. 3, no. 5, Jun. 2022, doi: 10.1007/s42979-022-01226-6.

[8] D. Swain *et al.*, 'Deep Learning Models for Yoga Pose Monitoring', *Algorithms*, vol. 15, no. 11, Art. no. 11, Nov. 2022, doi: 10.3390/a15110403.

[9] V. Anand Thoutam *et al.*, 'Yoga Pose Estimation and Feedback Generation Using Deep Learning', *Comput. Intell. Neurosci.*, vol. 2022, p. e4311350, Mar. 2022, doi: 10.1155/2022/4311350.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:42:56 UTC from IEEE Xplore. Restrictions apply.

[10] R. Jadhav, V. Ligde, R. Malpani, P. Mane, and S. Borkar, 'Aasna: Kinematic Yoga Posture Detection And Correction System Using CNN', *ITM Web Conf.*, vol. 56, p. 05007, 2023, doi: 10.1051/itmconf/20235605007.

[11] J. D. R, S. P. H, M. V, K. H. R, and P. R, 'Yoga Pose Estimation using Artificial Intelligence', in *2023 International Conference on Data Science and Network Security (ICDSNS)*, Jul. 2023, pp. 1–6. doi: 10.1109/ICDSNS58469.2023.10245467.

[12] D. Parashar, O. Mishra, K. Sharma, and A. Kukker, 'Improved Yoga Pose Detection Using MediaPipe and MoveNet in a Deep Learning Model', *Rev. Intell. Artif.*, vol. 37, no. 5, pp. 1197–1202, Oct. 2023, doi: 10.18280/ria.370511.

[13] F. B. Ashraf, M. U. Islam, M. R. Kabir, and J. Uddin, 'YoNet: A Neural Network for Yoga Pose Classification', *SN Comput. Sci.*, vol. 4, no. 2, p. 198, Feb. 2023, doi: 10.1007/s42979-022-01618-8.

[14] 'Yoga Pose Image classification dataset'. Accessed: Dec. 09, 2023. [Online]. Available: https://www.kaggle.com/datasets/shrutisaxena/yoga-pose-image-classification-dataset

[15] S. Garg, A. Saxena, and R. Gupta, 'Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application', *J. Ambient Intell. Humaniz. Comput.*, Jun. 2022, doi: 10.1007/s12652-022-03910-0.

[16] J. Z. Tan, C. P. Lee, K. M. Lim, and J. Y. Lim, 'Yoga Pose Estimation with Machine Learning', in *2023 11th International Conference on Information and Communication Technology (ICoICT)*, Aug. 2023, pp. 260–265. doi: 10.1109/ICoICT58202.2023.10262445.

[17] E. W. Trejo and P. Yuan, 'Recognition of Yoga Poses Through an Interactive System with Kinect Device', in *2018 2nd International Conference on Robotics and Automation Sciences (ICRAS)*, Jun. 2018, pp. 1–5. doi: 10.1109/ICRAS.2018.8443267.

[18] J. Palanimeera and K. Ponmozhi, 'Classification of yoga pose using machine learning techniques', *Mater. Today Proc.*, vol. 37, pp. 2930–2933, Jan. 2021, doi: 10.1016/j.matpr.2020.08.700.

[19] D. Shah, V. Rautela, C. Sharma, and A. Florence A, 'Yoga Pose Detection Using Posenet and k-NN', in *2021 International Conference on Computing, Communication and Green Engineering (CCGE)*, Pune, India: IEEE, Sep. 2021, pp. 1–4. doi: 10.1109/CCGE50943.2021.9776451.

[20] J. Guo *et al.*, 'An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring', *Comput. Netw.*, vol. 151, pp. 166–180, Mar. 2019, doi: 10.1016/j.comnet.2019.01.026.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:42:56 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575537

Assessing Gender Bias in Machine Translation

Rishabh Jain
Computer Science and Engineering
Delhi Technological University
Delhi, India
jrishabh720@gmail.com

***Abstract*—Machine translation systems, such as Google Translate and Bing Microsoft Translator, are widely utilized for translating text from one language to another. This study undertakes a comprehensive evaluation of Google Translate and Bing Microsoft Translator with a specific focus on natural gender translation. Through rigorous analysis, it assesses the occurrence of female, male, and neutral forms in the translations of personality adjectives, professions, and nouns. Emphasizing the English to Hindi language pair, the evaluation meticulously examines how effectively these translation systems handle gendered language nuances, particularly in professions where biases may exist. By providing statistical insights into the translation outputs, the study aims to shed light on potential biases and accuracy levels within these machine translation systems. Ultimately, this evaluation contributes to a deeper understanding of the capabilities and limitations of automated translation tools in accurately representing gender-specific language constructs, thereby facilitating more nuanced and sensitive cross-lingual communication.**

***Index Terms*—Gender bias, Machine Translation, N-grams, Natural Language Processing (NLP)**

I. Introduction

The growing connectivity of our world has made the ability to communicate across linguistic boundaries more crucial than ever. Language diversity poses a formidable barrier to connectivity. With the increasing interconnectedness through trade, travel, and digital communication, there is a burgeoning demand for machine translation that is both unbiased and accurate. Machine Translation (MT) Systems are important for breaking down language barriers when we talk to people worldwide. For translating text or speech from one language to another language these system use advance algorithms and NLP(Natural Language Processing) techniques. Although the development of reliable machine translators is challenging, MT systems are continuously improving and expanding their functionalities. They are constantly evolving and getting closer to achieving translation accuracy comparable to that of human expertise.

MT systems can show gender bias by favoring one gender over another based on professions, pronouns, and verbs. Mitigating these biases enhances fairness and equity, resulting in better user experience and equal treatment based on gender.

This study evaluates the gender bias in Machine Translation (MT) systems. When translating from a gender-neutral language like Hindi to a language with grammatical gender like English, which manifests in various ways across different languages. Languages classify words as masculine, feminine, or neutral in grammatical gender which varies between Hindi to English. For example, ‘doctor’ or ‘nurse’ might be perceived as typically male or female roles. It reflects both the real-world makeup of the data and the biases found in everyday communication. Sentence-by-sentence translation is a common feature of most modern MT systems. For managing the gender aspect in translation, these systems depend on the patterns they learned during training, but it becomes challenging when it comes to translating from a gender-neutral language like Hindi to a language with grammatical gender like English. A language with grammatical genders requires additional information to translate gender accurately. MT systems can be biassed based on recent studies but very little work is done in the Hindi language. For example, Google Translate favors masculine translations [1]. It doesn’t matter whether the source text is gender-neutral or not. There seems a bias in the machine translation system. To address this issue, Systematically evaluating bias in machine translation systems, the study concludes by comparing the original English sentences with their translated counterparts to identify any gender bias introduced during translation and to observe the influence of professions on bias.



Fig. 1. Occupation gender bias shown in some HI- EN (Hindi-English) translation systems

II. Related Work

Recent literature on bias in Natural Language Processing (NLP) has garnered attention, particularly focusing on con-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:38:07 UTC from IEEE Xplore. Restrictions apply.

trolling gender fluctuations in translations by major Machine Translation (MT) providers, with neural approaches gaining prominence. Rabinovich [2] conducted extensive experiments akin to Bawden [3], emphasizing gender preservation in PB-SMT systems through domain adaptation, treating female and male genders as separate domains. In Neural Machine Translation (NMT), Vanmassenhove [4] experimented with inserting artificial tokens at the sentence start to denote speaker gender, akin to Sennrich [5], who introduced 'informal' or 'polite' tags to training sentences. Elaraby [6] introduced a technique for speech-like text translation, particularly English-to-Arabic, training on generic data and fine-tuning with gender-labeled sentences for correct gender agreement. Moryossef [7] proposed a black-box method to control NMT translations in cases of gender ambiguity by concatenating unambiguous artificial antecedents with speaker and interlocutor information. Recent studies, recognizing the scarcity of publicly available corpora, designed corpora specifically to evaluate MT performance in gender translation (Font and Costa-Jussa [8]; Di Gangi [9]). Monti [10] provided an overview of gender-related issues in MT, while Sun [11] conducted a literature review on gender bias in NLP. Social bias within language models arises from unintended associations in data. Previous works addressing bias include Bolukbasi [12], employing gender-definition pairs and principal component analysis for bias direction inference, but Gonen and Goldberg [13] noted incomplete debiasing. Huang [14] measured bias by evaluating if sentiment changed with entity alterations, while Kurita [15] introduced a metric based on WEAT to measure BERT bias. Stanovsky devised a method to automatically evaluate gender bias in 8 languages, aligning translations with source text and extracting gender. Work on mitigating bias has primarily focused on English, despite being non-gendered. Zhou and Gonen [16] measured bias in Spanish and French, considering animate and inanimate nouns, and neutralizing gender signals in training, respectively. Research in quantifying and mitigating bias in NLP is extensive but lacks emphasis on Hindi and other Indic languages, possibly due to resource constraints. Pujari [17] pioneered geometric debiasing, defining a bias subspace to decompose words into gendered and neutral components, later classified using SVMs. Research efforts address gender bias in MT and language models through domain adaptation, token insertion, fine-tuning, black-box methods, and corpus design. Despite progress, mitigating bias in Indic languages remains underexplored, highlighting the need for further research in this area.

III. Methodology

A. Dataset

The dataset used in this study comprises 3888 sentences collected from https://github.com/argentina-res/gender_project and generated using Large Language Models (LLMs) tools. Each sentence is tagged with one of three genders: male (1821 sentences), female (1814 sentences), or neutral (253 sentences). This balanced dataset ensures fair evaluation of bias, as each sentence contains only one gender. Additionally, the dataset includes Hindi translations of the sentences, along with their respective gender labels.

B. Description of the MT Systems

Machine translation is a computerized process utilizing statistical or neural machine learning models to convert text from a source language to a target language.

1) Google Translator (GT): GT, launched in 2003 as a statistical MT system and transitioning to a neural system in 2016, provides translations at the sentence level. Since 2018, GT offers alternatives for ambiguous or under-specified English words in certain languages with male-female alternatives. However, this feature is not available for Hindi.

2) Bing Microsoft Translator (BMT): BMT is a machine translation system that is owned by Microsoft. Initially, it used a statistical approach (Monti, 2017), but later it switched to a neural system (Alma-hasees, 2018). Unlike GT, BMT does not offer alternatives in the translation box itself [18].

After generating translations using Google Translator and Bing Microsoft Translator, an analysis was conducted to determine the frequency of occurrences of female, male, and neutral forms. The focus was on gender-specific terms such as personality adjectives, professions, and nouns.

C. Bias Statement

Bias refers to the tendency to show preference or inequity towards certain groups. In the context of machine translation, bias can lead to unfair or inaccurate representations of different genders or identities. Biased translations may perpetuate stereotypes and exclude certain groups from the conversation. Addressing bias in machine translation is crucial for creating more inclusive and equitable representations that benefit all individuals and promote diversity and respect.

D. Detection of gender

The detection of gender in a given text can be accomplished by using N-grams (Uni-grams, Bi-grams, Tri-grams) and Transformers. However, this method may produce comparable outcomes for both the original English sentences and their translations, as the sentences include only one gender. The process begins with the tokenization of text during the preprocessing stage, which involves dividing it into individual words (tokens) for both the English and its translated versions. Then converting all the tokens to lowercase. This ensures that the analysis is completely insensitive to case. This stage also includes stemming or lemmatization techniques, which are used to normalize the words to their root form. The gender detection algorithm identifies words such as 'He', 'His', 'Him', and 'Himself', which indicate male gender, and cross-references them with pre-defined gender-specific lists. Similarly, for the female gender, words like 'She', 'Her', 'Hers', and 'Herself' are checked, while for the neutral gender, terms like 'They', 'We', 'Our', and 'Themselves' are considered. Depending on

[0]Code available at: https://github.com/jrishabh720/GenderBias

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:38:07 UTC from IEEE Xplore. Restrictions apply.

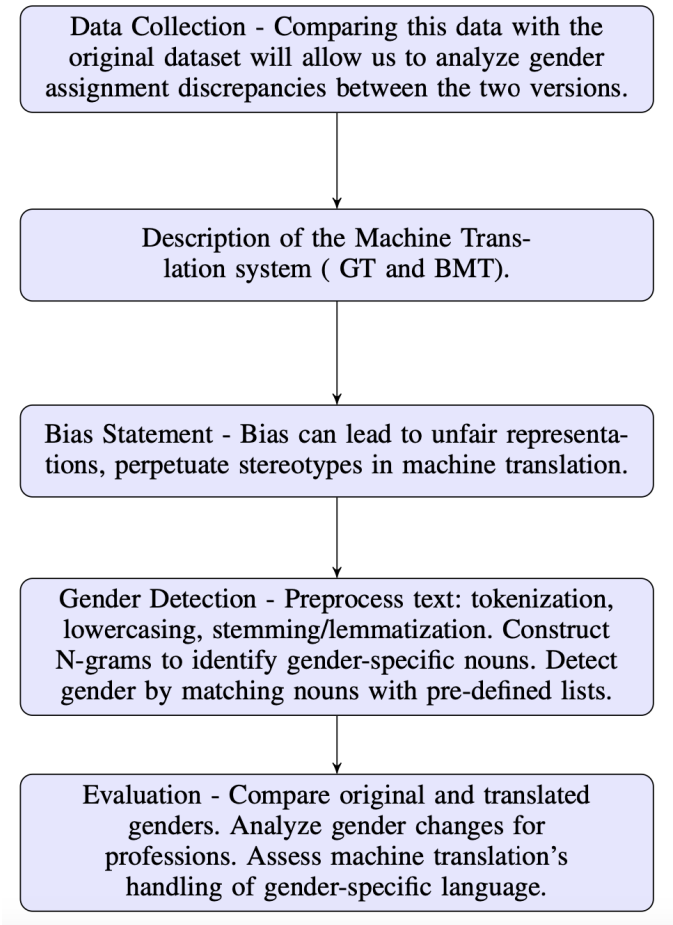


Fig. 2. Flow Diagram of Methodology

the match found in these lists, a gender label is assigned, categorizing the noun as ”male,” ”female,” or ”neutral.”

E. Evaluation

To evaluate the genders of the original data, comparisons are made with the genders of the translated data generated by Google Translator and Microsoft Bing Translator. Using various combinations, including comparing the results of the Original and Google Translator, Original and Bing Microsoft Translator, and Original, Google Translator, and Bing Microsoft Translator, patterns and correlations between professions and gender changes during translation are determined. The Natural Language Toolkit (NLTK) is utilized to extract professions from the dataset. This analysis aims to provide insights into how machine translation systems handle gender-specific language and its significance for professions during translation.

IV. Results

In this section, I intend to present the results and subsequent evaluation of the study. A comprehensive analysis and concrete examples will be provided. It is important to note that all evaluations were conducted in April 2024.

TABLE I
Gender Distribution in Dataset.

Data Gender	Value Count
Male	1821
Female	1814
Neutral	253

Table I distribution guarantees a fair evaluation across gender categories. The table provides a visual representation of these counts for easy comparison and analysis of gender distribution within the dataset.

TABLE II
Gender Distribution After Google Translation.

Google Translated Gender	Value Count
Male	2976
Female	386
Neutral	526

TABLE III
Gender Distribution After Bing Microsoft Translation.

BMT Gender	Value Count
Male	2946
Female	543
Neutral	399

Table II shows the Google translation's gender distribution across sentences. This provides insight into how gender-specific data is interpreted and categorized by the Google translation algorithm. By comparing this distribution with the original dataset, we can examine any disparities or similarities in gender assignment between the two versions and Table III shows the gender distribution of sentences translated by Bing Microsoft Translator.

TABLE IV
Gender Distribution in Original and Translated Data using Google Translator.

Data Gender - Google Translated Gender	Value Count
MM	1587
FM	1249
FF	361
MN	209
FN	204
NM	140
NN	113
MF	25

Table IV displays the gender distribution across sentences in the original dataset and the translated version using Google Translator. The table presents counts for different gender combinations. MF represents cases where the initial gender was male but was categorized as female after translation. This comparison identifies any shifts in gender categorization

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:38:07 UTC from IEEE Xplore. Restrictions apply.

TABLE V
GENDER DISTRIBUTION IN ORIGINAL AND TRANSLATED DATA USING BMT. FM REPRESENTS DATA WHERE THE INITIAL GENDER WAS FEMALE (F) AND AFTER TRANSLATION, IT WAS CATEGORIZED AS MALE (M).

Data Gender - BMT Gender	Value Count
MM	1627
FM	1273
FF	465
NN	207
MN	116
MF	78
FN	76
NM	46

TABLE VI
GENDER DISTRIBUTION AND COMPARISON IN DATA, GOOGLE TRANSLATED, AND BMT.

Result	Value Count
MMM	1464
FMM	1113
FFF	332
MNM	145
FNM	134
FMF	112
NNN	107
NMN	100
MMF	67
MNN	60
MMN	56
FNN	49
NMM	40
FFM	26
FMN	24
FNF	21
MFM	18
MFF	7
NNM	6
MNF	4
FFN	3

between the original and translated data, providing insights into the impact of translation on gender assignment. Google Translator often fails to appropriately translate gender-specific terms which can result in misclassification. For example, neutral roles may be erroneously rendered as male-dominated or vice versa. Google Translator may perpetuate gender biases and stereotypes by assigning incorrect gender to certain professions in certain languages. For instance, it may assign a male gender to gender-neutral professions like 'Mechanic,' 'Developer,' 'Doctor,' and 'Guard' in certain English languages. Similarly, it may assign a female gender to professions like 'Nurse,' 'Veterinarian,' 'Teacher,' when translating from Hindi to English where gendered pronouns are prevalent. Table IV and Table V show that the gender distribution in the original dataset and the translated versions using Bing Microsoft Translator and Google Translator.

Table VI shows the gender distribution and comparison

Original	Hindi	GT	BMT
The guard appreciated getting treatment from the nurse because he did a good job. (M)	गार्ड ने नर्स से इलाज करने की सराहना की क्योंकि उसने अच्छा काम किया। (N)	The guard appreciated treating the nurse as she did a good job.(F)	The guard appreciated the treatment from the nurse because she did a good job.(F)
The physician told the farmer that she tried the best.(F)	चिकित्सक ने किसान को बताया कि उसने सबसे अच्छी कोशिश की। (N)	The doctor told farmer that he tried his best.(M)	The therapist told farmer that he tried the best.(M)
The manager told the hairdresser that she expected something better.(F)	प्रबंधक ने हेयरड्रेसर को बताया कि उसे कुछ बेहतर होने की उम्मीद थी। (N)	The manager told the hairdresser that he expected something better.(M)	The manager told the hairdresser that he expected something better.(M)

Fig. 3. Example Sentences and Their Translations

across the original dataset, Google translated dataset, and Bing Microsoft Translator (BMT) translated dataset. Each row refers to a distinct combination of genders, with counts. The first character represents the gender of the original dataset, the second character represents the gender of the Google-translated dataset, and the third character represents the gender of the BMT-translated dataset. Counts in the table give an idea of gender assignment differences or similarities between the translation models. As one example, the combination of male-male-male, denoted by the code 'MMM,' has 1464, indicating that classify a sentence as male. Such comparative study will provide some insight into how rightly or wrongly different translation models translate gender and shed some light on the introduction of discrepancies or bias during the translation process.

The study examined the gender translation performance of two leading models of translation across a range of professions. The models, though very advanced in their algorithms, in the translation process based on profession convert gender from female to male or male to female. These are indications that improvements are needed in automated translation systems. Enhancing the accuracy of gender translation is crucial for promoting fair representation and inclusivity across linguistic contexts.

In this study, accuracy is defined as the absence of bias in the translated text compared to the original data. To calculate the accuracy, the counts of correctly translated sentences are summed and then divided by the total number of sentences in the dataset (3888 in this case).

For the Google translation system and Bing Microsoft translation system, the accuracy is calculated as follows:

$$\text{Accuracy (Google)} = \frac{1587 + 361 + 113}{3888} \approx 53.03\%$$

$$\text{Accuracy (Bing)} = \frac{1627 + 465 + 207}{3888} \approx 59.12\%$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:38:07 UTC from IEEE Xplore. Restrictions apply.

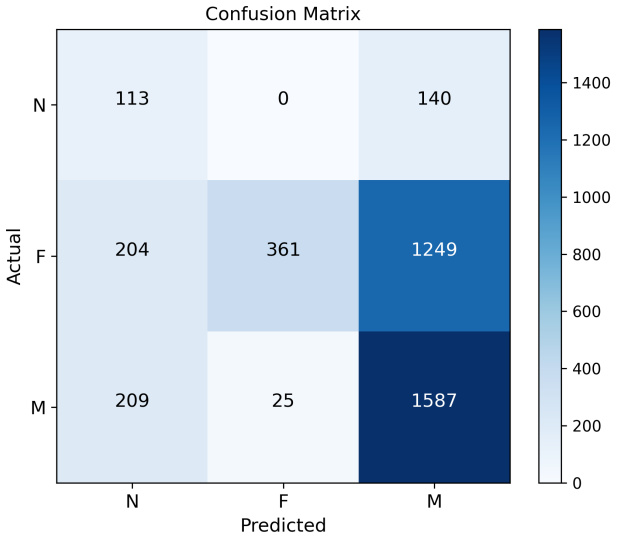


Fig. 4. Confusion Matrix for Google Translator(GT)

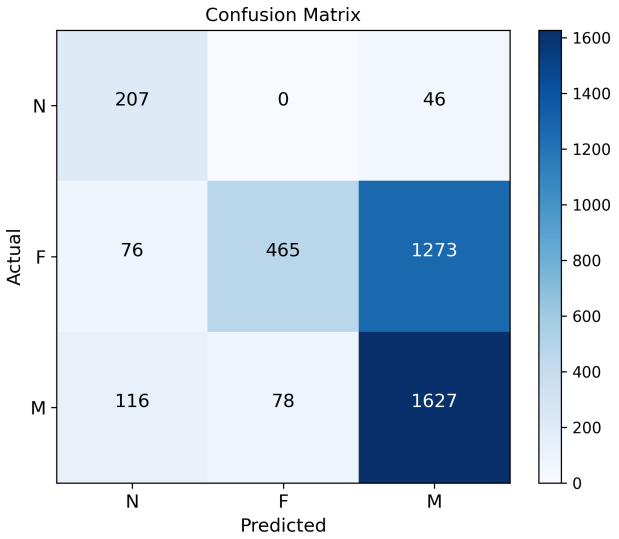


Fig. 5. Confusion Matrix for Bing Microsoft Translator

Comparing the two, it's observed that the Bing Microsoft translation system achieves a slightly higher accuracy compared to Google. However, it's important to note that both systems exhibit similar performance in accurately translating sentences without introducing bias.

V. Conclusion and Future Work

In this study aimed to evaluate the gender translation performance of two major translation models across different professions. Despite their advanced algorithms, both systems struggled to accurately translate professions and often misinterpreted gender. The accuracy of Bing Microsoft and Google translation systems was compared by assessing their ability to translate sentences without introducing bias. Although Bing Microsoft showed slightly higher accuracy, both systems performed similarly in this regard.

In future work, the aim is to expand the evaluation to include a broader range of language pairs and a more diverse vocabulary. Additionally, plans are in place to develop a challenge set that specifically targets gender-related phenomena in language, enabling automatic evaluation of translation systems. The goal is to train a state-of-the-art machine translation (MT) system to investigate the impact of machine bias on translation outputs and explore potential mitigation strategies. Various approaches will be explored, drawing on methods proposed for other languages, to address and reduce biases in machine translation. Furthermore, accuracy can be improved by mitigating bias through gender tagging at the sentence or word level, thereby enhancing the performance and reliability of the translator.

References

[1] K. Ramesh, G. Gupta, and S. Singh, "Evaluating gender bias in Hindi-English machine translation," in *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, M. Costa-jussa, H. Gonen, C. Hardmeier, and K. Webster, Eds. Online: Association for Computational Linguistics, Aug. 2021, pp. 16–23. [Online]. Available: https://aclanthology.org/2021.gebnlp-1.3

[2] E. Rabinovich, R. N. Patel, S. Mirkin, L. Specia, and S. Wintner, "Personalized machine translation: Preserving original author traits," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, M. Lapata, P. Blunsom, and A. Koller, Eds. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 1074–1084. [Online]. Available: https://aclanthology.org/E17-1101

[3] R. Bawden, G. Wisniewski, and H. Maynard, "Investigating gender adaptation for speech translation," in *Actes de la conférence conjointe JEP-TALN-RECITAL 2016. volume 2 : TALN (Posters)*, L. Danlos and T. Hamon, Eds. Paris, France: AFCP - ATALA, 7 2016, pp. 490–497. [Online]. Available: https://aclanthology.org/2016.jeptalnrecital-poster.23

[4] E. Vanmassenhove, C. Hardmeier, and A. Way, "Getting gender right in neural machine translation," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, Eds. Brussels, Belgium: Association for Computational Linguistics, Oct.-Nov. 2018, pp. 3003–3008. [Online]. Available: https://aclanthology.org/D18-1334

[5] R. Sennrich, "How grammatical is character-level neural machine translation? assessing MT quality with contrastive translation pairs," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, M. Lapata, P. Blunsom, and A. Koller, Eds. Valencia, Spain: Association for Computational Linguistics, Apr. 2017, pp. 376–382. [Online]. Available: https://aclanthology.org/E17-2060

[6] M. Elaraby, A. Y. Tawfik, M. Khaled, H. Hassan, and A. Osama, "Gender aware spoken language translation applied to english-arabic," 2018.

[7] A. Moryossef, R. Aharoni, and Y. Goldberg, "Filling gender number gaps in neural machine translation with black-box context injection," 2019.

[8] M. R. Costa-Jussà, "An analysis of gender bias studies in natural language processing," *Nature machine intelligence*, vol. 1, no. 11, pp. 495–496, 10 2019. [Online]. Available: https://doi.org/10.1038/s42256-019-0105-5

[9] L. Bentivogli, B. Savoldi, M. Negri, M. A. Di Gangi, R. Cattoni, and M. Turchi, "Gender in danger? evaluating speech translation technology on the MuST-SHE corpus," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 6923–6933. [Online]. Available: https://aclanthology.org/2020.acl-main.619

[10] J. Monti, *Gender issues in machine translation*, 6 2020. [Online]. Available: https://doi.org/10.4324/9781315158938-39

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:38:07 UTC from IEEE Xplore. Restrictions apply.

[11] T. Sun, A. Gaut, S. Tang, Y. Huang, M. ElSherief, J. Zhao, D. Mirza, E. Belding, K.-W. Chang, and W. Y. Wang, "Mitigating gender bias in natural language processing: Literature review," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds. Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 1630–1640. [Online]. Available: https://aclanthology.org/P19-1159

[12] T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, and A. Kalai, "Man is to computer programmer as woman is to homemaker? debiasing word embeddings," 2016.

[13] M. Geva, Y. Goldberg, and J. Berant, "Are we modeling the task or the annotator? an investigation of annotator bias in natural language understanding datasets," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 1161–1166. [Online]. Available: https://aclanthology.org/D19-1107

[14] P.-S. Huang, H. Zhang, R. Jiang, R. Stanforth, J. Welbl, J. Rae, V. Maini, D. Yogatama, and P. Kohli, "Reducing sentiment bias in language models via counterfactual evaluation," in *Findings of the Association for Computational Linguistics: EMNLP 2020*, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 65–83. [Online]. Available: https://aclanthology.org/2020.findings-emnlp.7

[15] K. Kurita, N. Vyas, A. Pareek, A. W. Black, and Y. Tsvetkov, "Measuring bias in contextualized word representations," in *Proceedings of the First Workshop on Gender Bias in Natural Language Processing*, M. R. Costa-jussà, C. Hardmeier, W. Radford, and K. Webster, Eds. Florence, Italy: Association for Computational Linguistics, Aug. 2019, pp. 166–172. [Online]. Available: https://aclanthology.org/W19-3823

[16] H. Gonen and Y. Goldberg, "Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them," 2019.

[17] A. K. Pujari, A. Mittal, A. Padhi, A. Jain, M. K. Jadon, and V. Kumar, "Debiasing gender biased hindi words with word-embedding," *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:211104718

[18] A. A. Rescigno, J. Monti, A. Way, and E. Vanmassenhove, "A case study of natural gender phenomena in translation: A comparison of Google Translate, bing Microsoft translator and DeepL for English to Italian, French and Spanish," in *Workshop on the Impact of Machine Translation (iMpacT 2020)*, S. O'Brien and M. Simard, Eds. Virtual: Association for Machine Translation in the Americas, Oct. 2020, pp. 62–90. [Online]. Available: https://aclanthology.org/2020.amta-impact.4

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:38:07 UTC from IEEE Xplore. Restrictions apply.

Astrophysics and Space Science (2024) 369:73
https://doi.org/10.1007/s10509-024-04338-y

RESEARCH

Barrow holographic dark energy model in $f(R,T)$ theory

Kanchan Devi[1] · **Ajay Kumar**[2] · **Pankaj Kumar**[1]

Received: 29 February 2024 / Accepted: 16 July 2024 / Published online: 26 July 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

We consider a recently modified version of holographic dark energy, known as Barrow holographic dark energy, in the framework of $f(R,T)$ gravity and consider a flat Friedmann-Lemaitre-Robertson-Walker line element for our study. We solve the field equations of the model to obtain the values of Hubble parameter and scale factor of the universe. We obtain the values of deceleration parameter and effective equation of state to discuss the evolution of the universe. Further, we constrain the model parameters using various data sets like type Ia supernova, observational Hubble data, SH0ES data etc. We use the Monte Carlo Markov Chain method to obtain the best fit values of the model parameters. We observe that the best fit present values of Hubble parameter $H_0 = 67.764^{+1.274}_{-1.354}$ and $H_0 = 70.440^{+0.816}_{-0.869}$, obtained for two different combinations of observational data, are in agreement with recent observations. We also constrain the case in which our model converts to the Barrow holographic dark energy model in general relativity and compare the results of both models. We compare the results with Λ-CDM model wherever required. We plot deceleration parameter against redshift parameter for best fit values of the model parameters. We observe a smooth phase transition of the universe from early time decelerated expansion to accelerated expansion which shows compatibility with recent observations. The values of equation of state parameter ω_h of Barrow holographic dark energy are found to be $-0.873^{+0.115}_{-0.078}$ and $-0.866^{+0.156}_{-0.130}$, and age of the universe are found to be 14.09 Gyr and 15.43 Gyr for two combination of data sets for Barrow holographic dark energy model in $f(R,T)$ gravity. Furthermore, we apply statefinder and Om diagnostic to discriminate our model from existing dark energy models.

Keywords Barrow holographic dark energy · $f(R,T)$ theory · Accelerated expansion · Observational estimation · Statefinder analysis · Om diagnostic

1 Introduction

The cosmological observations have changed our understanding about the universe in a significant way specially the observations of last three decades have made overwhelming changes in our knowledge of the universe. In the year 1998, the observations of type Ia supernovae changed our understanding about cosmic evolution. The accelerated expansion of the universe was observed by Riess et al. (1998) and Perlmutter et al. (1999), and this observational discovery presented a challenge in front of cosmologists to explain it theoretically. The observations of various cosmological probes like Wilkinson Microwave Anisotropy Probe (WMAP) (Spergel et al. 2003, 2007), Sloan Digital Sky Survey (SDSS) (Tegmark et al. 2004a, 2004b) and Planck data (Ade et al. 2014; Aghanim et al. 2020) etc. have confirmed the accelerated expansion of the universe. These observations suggest a mysterious type of energy, the so called dark energy (DE), which exerts a negative pressure required to accelerate the universe. According to recent observations, the universe should consist about 68.3% DE, 26.8% cold dark matter (CDM), and only 4.9% usual baryonic matter which can be described by the well-known particle theory (Aghanim et al. 2020). Till now, DE is a hypothetical form of energy because no evidences of it's true nature and origin are available. It behaves opposite of gravity and supposed to be the energy of the empty space. The simplest DE model

✉ P. Kumar
pankaj.11dtu@gmail.com

K. Devi
kanchanmehta64@gmail.com

A. Kumar
maths.ajy@gmail.com

[1] Department of Mathematics, Maharaja Agrasen University, Atal Shiksha Kunj, Kalujhanda, Solan, Himachal Pradesh, 174103, India

[2] Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi, 110042, India

that fits the cosmological observations in a significant way is the Λ-CDM model, where Λ is the cosmological constant (Carroll 2001; Peebles and Ratra 2003). However, a number of DE models have been proposed in the literature but the Λ-CDM model is the most successful model of DE till date and known as the standard model of cosmic evolution. However, this model is not free from problems and the problems it is facing are cosmic coincidence, fine-tuning and H_0 tension (Copeland et al. 2006; Sahni et al. 2014; Valentino et al. 2021a,b).

The second possibility to explain accelerated expansion of the universe is by assuming that the General Relativity (GR) breaks down at large scales and a more general action is required to describe the gravitational field. The modified theories of gravity have been obtained by modifying the Einstein-Hilbert (EH) action of GR. The study on evolution of the universe through modified theories of gravity have become a popular area of research in recent years. The $f(R)$ theory is considered one of the simplest modifications of GR in which the gravitational Lagrangian depends on a general function of Ricci curvature scalar R (Bamba et al. 2008; Nojiri and S.D. Odintsov 2011; Matos et al. 2021). Buchdahl (1970) was first who considered the general function $f(R)$ in the EH action to present a more general model of gravity. It is natural to look beyond the dependence of gravitational Lagrangian on a function of R alone, it might be a function of two or more terms. Considering the fact, a number of theories such as $f(R, G)$ gravity (Bamba et al. 2010; Elizalde et al. 2010; Odintsov et al. 2019), $f(R, T)$ gravity (Singh and Kumar 2014), $f(Q, T)$ gravity (Xu et al. 2019), $f(T, B)$ gravity (Bahamonde et al. 2015; Wright 2016) etc. have been proposed to explain the recent accelerated expansion and other phenomena of the universe. Clifton et al. (2012) have provided a detailed review on modified theories of gravity.

In the present work, we are interested in $f(R, T)$ gravity which was proposed by Harko et al. (2011), where R as usual stands for Ricci curvature scalar and T denotes the trace of energy-momentum tensor. The dependence on T may be induced by exotic imperfect fluids or quantum effects (conformal anomaly). The authors have shown that a particular class of models in $f(R, T)$ gravity may produce an effective cosmological constant $\Lambda(t)$. Thus, the $f(R, T)$ gravity may provide a richer cosmological theory of gravity which also provides the space for imperfect fluids and quantum effects in its formalism. In a very recent paper, Bouali et al. (2023) have shown that a cosmological model in scalar-tensor $f(R, T)$ gravity is better than the Λ-CDM model in the light of recent observations. A number of authors have discussed the $f(R, T)$ gravity in different context to explain the early and late time evolution of the universe (Nojiri and Odintsov 2011; Houndjo 2012; Naser and Pradhan 2014). The reconstruction of $f(R, T)$ gravity has been carried out by Singh and Singh (2014) for two well-known scale factors. The cosmological implications of bulk viscosity and particle creation in the framework of $f(R, T)$ gravity was studied by Singh and Kumar (2014) and Kumar and Singh (2015). Sahoo et al. (2016) studied the Kaluza Klein cosmology in $f(R, T)$ theory. The holographic Ricci dark energy and generalized second law of thermodynamics has been discussed by Singh and Kumar (2018) within the framework of $f(R, T)$ gravity. Aktas (2019) has described various DE models in $f(R, T)$ theory like the k-essence, tachyon field. In recent papers, Bhattacharjee et al. (2020) have presented an inflationary model in $f(R, T)$ theory, energy conditions in unimodular $f(R, T)$ gravity have been discussed in (Rajabia and Nozarib 2021) and observational constraints have been obtained on $f(R, T)$ gravity using a particular form $f(R, T) = R + h(T)$ by Sardar et al. (2023).

To have a more accurate and general theory of gravity, the quantum physics should be included in the theory of gravity. Some efforts have been made but still we are far from the accurate theory of quantum gravity. Some DE candidates have been proposed in the literature which contains some properties of quantum gravity. A particularly interesting DE candidate featuring the properties of quantum gravity is the holographic dark energy (HDE). As the name suggests, this model is based on the holographic principle which was proposed by 't Hooft (1993). The energy density $\rho_h = 3c^2 M_p^2 L^{-2}$ of the HDE can be described by only two physical quantities which are the Planck mass M_p and the cosmic length scale L (Li 2004). Saridakis (2020) proposed a general version of HDE using the Barrow entropy instead of the standard Bekenstein-Hawking one which is known as Barrow holographic dark energy (BHDE). In a recent paper (Anagnostopoulos et al. 2020), observational constraints have been found on BHDE in flat FLRW universe taking GR as theory of evolution. Some recent works on BHDE are given in refs. (Srivastava and Sharma 2020; Mamon et al. 2021; Adhikary et al. 2021; Oliveros et al. 2022; Paul et al. 2022).

Singh and Kumar (2016) have discussed HDE in the framework of $f(R, T)$ gravity and applied statefinder analysis to the model. Saridakis (2020) found that BHDE exhibits more interesting and richer phenomenology comparing to the standard HDE scenario, therefore, it is worthy to study BHDE in $f(R, T)$ gravity. In the case of HDE, Hsu (2004) and Li (2004) have shown that Hubble horizon as an IR cut-off leads to a wrong equation of state, therefore, the accelerated expansion can not be achieved in the formalism of GR. Later on, Pavón and Zimdahl (2005) showed that interacting HDE may solve the problem and we may explain the accelerated expansion of the universe. Further, it was shown in a number of papers (Zimdahl and Pavón 2007; Campo et al. 2011; Singh and Kumar 2017) that the Hubble horizon as an IR cut-off is a suitable candidate to explain the recent accelerated expansion and it is also free from causality problem.

Moreover, it has been shown that Hubble horizon as an IR cut-off may explain the accelerated expansion of the universe without interaction in the formalism of $f(R,T)$ gravity (Singh and Kumar 2016). Therefore, we consider Hubble horizon as an IR cut-off in the present model. Tsallis HDE, generalization of HDE using Tsallis entropy, has also been discussed in $f(R,T)$ gravity. The authors studied Tsallis HDE for various IR cut-off and observed instability in the models (Zubair and Durrani 2021).

Sahni et al. (2003) proposed a new geometrical diagnostic pair {r, s} to discriminate among DE models called the statefinder. Further, Alam et al. (2003) discussed the diagnostic pair {r, s} in details. The statefinder is a geometrical diagnostic and allows us to characterize the properties of DE in a model-independent manner. The statefinder is dimensionless and is constructed from the scale factor of the cosmic expansion and its time derivatives only. The parameter r comes next in the hierarchy of geometrical cosmological parameters after the Hubble parameter H and the deceleration parameter q, while s is a linear combination of q and r. The ΛCDM and standard cold dark matter (SCDM) models have fixed values $\{1, 0\}$ and $\{1, 1\}$ of the statefinder pair {r, s}, respectively. Another geometrical diagnostic is the Om diagnostic which was proposed by Sahni et al. (2008). This diagnostic also applies to discriminate among DE models. We can discriminate among DE models on the basis of positive and negative slope of trajectories in $Om(z)$-z plane.

The main motive of the paper is to discuss BHDE model in the framework of $f(R,T)$ gravity whereas HDE model has been discussed in the theory. In the framework of GR, the BHDE model has been discussed by several authors showing the recent accelerated expansion successfully (Srivastava and Sharma 2020; Adhikary et al. 2021; Oliveros et al. 2022; Paul et al. 2022). But, problems like H_0 tension remain intact with this DE model also in GR as it can be observed in an observational analysis done in the paper (Anagnostopoulos et al. 2020). As we have discussed above $f(R,T)$ gravity presents a richer theory of gravity than GR, therefore, it is important to discuss the BHDE model in the framework of $f(R,T)$ gravity. It will be interesting to see whether the BHDE model in $f(R,T)$ gravity may resolve problems like H_0 tension, age problem, etc or not. In the era of observational precision, the cosmological model should be tested in light of recent observations. Therefore, we apply Monte Carlo Markov Chain (MCMC) method to find the best fit values of model parameters. Then, we use the best fit values to discuss the evolution of the universe within our model using cosmological parameters. Moreover, we also constrain our model for the case in which it converts into the BHDE model in GR. However, the BHDE model in GR has been constrained in the paper (Anagnostopoulos et al. 2020) but we use the latest observational data to constrain it. Then, we compare the results of our model with the results of this model. We also apply statefinder and Om diagnostic to our model to discuss similarities and differences between our model and existing DE models.

The paper is organized as follows. Section 2 is devoted to model and field equations. In Sect. 3, we provide the solution of the field equations where we obtain the values of Hubble parameter H, cosmic scale factor a, deceleration parameter q and effective equation of state parameter ω_{eff}. In Sect. 4, we estimate the best fit values of the model parameters. The evolution of the universe is the topic of the Sect. 5 where we discuss the evolution of the universe through cosmological parameters using the best fit values of model parameter and use geometric diagnostics to compare our model with existing ones. Section 6 is devoted to the conclusion of the paper.

2 Model and field equations

The general form of the EH action for the $f(R,T)$ gravity is as follows (in the unit $8\pi G = 1$)

$$S = \frac{1}{2}\int [f(R,T) + 2\mathcal{L}_m]\sqrt{-g}d^4x, \tag{1}$$

where g stands for the determinant of the metric tensor $g_{\mu\nu}$, R denotes Ricci curvature scalar and $\mathcal{L}_m$ denotes the matter Lagrangian density. Here, T represents the trace of energy-momentum tensor, i.e $T = g^{\mu\nu}T_{\mu\nu}$, and energy-momentum tensor $T_{\mu\nu}$ can be obtained from matter Lagrangian density as

$$T_{\mu\nu} = \frac{-2\delta(\sqrt{-g}\mathcal{L}_m)}{\sqrt{-g}\delta g^{\mu\nu}}. \tag{2}$$

It has been considered in $f(R,T)$ gravity that $\mathcal{L}_m$ depends only on $g_{\mu\nu}$, not on its derivatives. Now, the above form of $T_{\mu\nu}$ simplifies to

$$T_{\mu\nu} = g_{\mu\nu}\mathcal{L}_m - 2\frac{\partial\mathcal{L}_m}{\partial g^{\mu\nu}}. \tag{3}$$

Varying the action (1) with respect to the metric tensor $g^{\mu\nu}$, we obtain the field equations of the theory as

$$f_R(R,T)R_{\mu\nu} - \frac{1}{2}f(R,T)g_{\mu\nu} + (g_{\mu\nu}\Box - \nabla_\mu\nabla_\nu)f_R(R,T) = T_{\mu\nu} - f_T(R,T)T_{\mu\nu} - f_T(R,T)\Theta_{\mu\nu}, \tag{4}$$

where f_R and f_T denote the derivatives of the functional $f(R,T)$ with respect to R and T, respectively. Here, $\Box$ denotes the d'Alembert operator and ∇_μ denotes the covariant derivative. The term Θ present in the field equation is given by

$$\Theta\mu\nu = -2T_{\mu\nu} + g_{\mu\nu}\mathcal{L}_m - 2g^{\alpha\beta}\frac{\partial^2\mathcal{L}_m}{\partial g^{\mu\nu}\partial g^{\alpha\beta}}. \tag{5}$$

A number of forms of the functional $f(R,T)$ can be used to obtain various classes of $f(R,T)$ gravity models. The popular forms of the functional $f(R,T)$ are $R+f(T)$, $f_1(R)+f_2(T)$, $f_1(R)+f_2(R)f_3(T)$ etc. which have been taken to discuss various aspects of cosmological evolution. In the present paper, we are interested in the form $f(R,T)=R+f(T)$, where the term $f(T)$ is modifying GR by including geometry and matter coupling. The Eq. (4) for $f(R,T)=R+f(T)$ get converted into following equation

$$R_{\mu\nu}-\frac{1}{2}Rg_{\mu\nu} = T_{\mu\nu}-(T_{\mu\nu}+\Theta_{\mu\nu})f'(T)+\frac{1}{2}f(T)g_{\mu\nu}, \tag{6}$$

where a prime denotes derivative with respect to the argument. The matter lagrangian $\mathcal{L}_m$ may be chosen as $\mathcal{L}_m=-p$, where p is the thermodynamic pressure of matter content of the universe. Now Eq. (5) gives $\Theta_{\mu\nu}=-2T_{\mu\nu}-pg_{\mu\nu}$. Using this result, Eq. (6) reduces to

$$R_{\mu\nu}-\frac{1}{2}Rg_{\mu\nu} = T_{\mu\nu}+(T_{\mu\nu}+pg_{\mu\nu})f'(T)+\frac{1}{2}f(T)g_{\mu\nu}. \tag{7}$$

In $f(R,T)$ gravity, it is interesting to note that energy-momentum tensor does not conserve ($\nabla_v T_{\mu\nu}\neq 0$) which leads to some interesting cosmological consequences. One of the important consequences is the particle production through an open system irreversible process from the gravitational field (Harko 2014). Recently, a scalar-tensor representation of $f(R,T)$ gravity has been proposed which reduces mathematical complexity of the theory and presents some interesting features (Rosa 2021; Rosa and Kull 2022). It has been shown in a recent paper (Pinto et al. 2022) that the scalar-tensor $f(R,T)$ gravity shows capabilities of explaining particle production and the entropy production, which can explain the accelerated expansion of the universe. But, in the present paper, we are not considering the possibility of irreversible particle production.

In the present work, we consider a spatially homogenous and isotropic flat Friedmann-Lemaitre-Robertson-Walker (FLRW) space-time which is expressed in co-moving coordinates by the line element

$$ds^2=dt^2-a^2(t)(dx^2+dy^2+dz^2). \tag{8}$$

For the line element (8), the Eq. (7) produces the following equations

$$3H^2=\rho+(\rho+p)f'(T)+\frac{1}{2}f(T), \tag{9}$$

$$2\dot{H}+3H^2=-p+\frac{1}{2}f(T), \tag{10}$$

where ρ and p denote total content density and pressure of the universe, respectively. We assume $f(T)=\alpha T$, where α is a coupling parameter. Thus, the functional has the form $f(R,T)=R+\alpha T$ which represents a linear coupling between matter and geometry. This linear functional is important because it can produce effective cosmological constant model as a special case. Further, it may provide additional quantum corrections required to explain the problems like coincidence problem and H_0 tension associated with the concordance model. For this linear functional, the field equations (9) and (10) reduce as

$$3H^2=\rho+\alpha(\rho+p)+\frac{1}{2}\alpha T, \tag{11}$$

$$2\dot{H}+3H^2=-p+\frac{1}{2}\alpha T. \tag{12}$$

Now, the matter aries to choose the components of energy budget of the universe. However, the modified theories were considered to provide an alternative to DE still there are a number of DE candidates which naturally exit in the universe like holographic dark energy (HDE) and logical to be considered in the framework of modified theories of gravity. Moreover, the evolution of the universe is a very complex phenomenon and it is very possible that it may not be explained entirely by one theory. Taking into consideration of this fact, a number of DE candidates have been studies in the framework of modified theories of gravity (Sheykhi 2009; Sharif and Nawazish 2017; Samart et al. 2021; Zubair and Durrani 2021; Pinki and Kumar 2023). The HDE is based on holographic principle, contains some significant properties of quantum gravity and it appears from bound on zero point energy coming from black hole formation. Saridakis (2020) proposed a general version of HDE, using the Barrow entropy instead of the standard Bekenstein-Hawking one, which is known as BHDE. Author found that BHDE exhibits more interesting and richer phenomenology comparing to the standard HDE scenario. Therefore, it is worthy to study BHDE model in $f(R,T)$ gravity, however, HDE models have already been discussed in this theory. The energy density of BHDE ρ_h is given by

$$\rho_h=3c^2H^{2-\Delta}, \tag{13}$$

where $0\leq\Delta\leq 1$. Here, we have considered Hubble horizon as IR cut-off which is the simplest choice and free from causality problem. Also, we have considered $M_p^2=1$. It is to be noted that for $\Delta=0$, we obtain the energy density of HDE.

We consider that the universe is filled with BHDE and Dark Matter (DM), and we don't consider baryonic matter due to it's negligible contribution to the late time evolution. Thus, total energy density is given by $\rho=\rho_h+\rho_m$, where ρ_h

and ρ_m denote density of BHDE and density of DM, respectively. As DM is a pressureless component, the total pressure is given by $p = p_h$, where p_h denotes pressure of BHDE. In the literature, the study of DE and DM has been considered in two ways, interacting or non-interacting. We can deal with the interaction between dark components through energy conservation equations. In general, the energy-momentum tensor does not conserve in $f(R,T)$ gravity, although, it conserves for particular classes of $f(R,T)$ gravity (Bertini and Velten 2023). One can easily discuss the interaction between dark components in those classes of $f(R,T)$ gravity where conservation holds. But, in the classes where it does not hold, the problem is that how to write the separate conservation equations of dark components like for BHDE and DM in the present case. However, the part avoiding conservation of energy-momentum tensor may be considered either in the conservation equation of BHDE or DM, or may be divided into conservation equations of both. So, it is quite possible to study the interaction between dark components in the classes of $f(R,T)$ gravity where the energy-momentum tensor does not conserve. There is no compelling evidence of interaction between dark components, therefore, we are not considering interaction between BHDE and DM in the present work. Now, the field equations (11) and (12) takes the form

$$3H^2 = \rho_h + \rho_m + \alpha(\rho_h + \rho_m + p_h) + \frac{1}{2}\alpha T, \tag{14}$$

$$2\dot{H} + 3H^2 = -p_h + \frac{1}{2}\alpha T. \tag{15}$$

The above two equations are independent field equations of our model. Here, we have mainly three unknown quantities H, ρ_m and p_h, and only two field equations. Therefore, we have to either assume a quantity or assume a relation between any two quantities to solve equations. In the next section, we assume a relation between ρ_h and p_h to solve equations.

3 Solution of field equations

Let us consider a barotropic equation of state (EoS) of BHDE given by $p_h = \omega_h \rho_h$, where ω_h denotes the EoS parameter. The barotropic EoS is a direct relationship between density ρ_h and pressure p_h which allows for analytical solutions of the field equations and makes the study of evolution of the universe easy. Furthermore, it provides a way to classify the type of fluids depending on the values of the EoS parameter ω_h. It makes also possible to explain the accelerated or decelerated expansion of the universe using the values of ω_h. Thus, barotropic EoS plays an important role in cosmological modelling of the universe. The trace of energy-momentum tensor is given by $T = \rho_m + \rho_h - 3p_h$. After combining (14) and (15), we obtain a single evolution equation of H as

$$2\dot{H} + (1+\alpha)[(1+\omega_h)\rho_h + \rho_m] = 0. \tag{16}$$

Let us solve Eq. (16) to obtain the value of H. To solve this equation, we require the value of ρ_m which can be obtained using (13) into (14) as

$$\rho_m = \frac{6H^2 - 3c^2(2+3\alpha-\alpha\omega_h)H^{2-\Delta}}{(3\alpha+2)}. \tag{17}$$

We observe that the matter density depends on the model parameters. If we observe the numerator term, which may be written as $6H^2 - 3c^2H^{2-\Delta}[2+\alpha(3-\omega_h)]$, it is easy to observe for $\alpha > 0$ that as we increase the value of α the value of the numerator term decreases. As α increases the value of denominator term $3\alpha+2$ increases. Thus, the large value of parameter α slows down the evolution of matter density ρ_m. This also shows that the evolution of matter density ρ_m in $f(R,T)$ gravity is slower than GR where $\alpha = 0$. For negative values of parameter α, such that we must observe $\rho_m > 0$, we observe the reverse effect on the evolution of ρ_m. In this case, the evolution of matter density ρ_m in $f(R,T)$ gravity is faster than GR. It is easy to observe that the small value of parameter Δ slows down the evolution of matter density ρ_m and vice-versa. Similarly, the large value of c also slows down the evolution of matter density ρ_m and vice-versa. Further, we observe for $\alpha > 0$ that a large negative value of EoS of BHDE ω_h slows down the evolution of matter density ρ_m and vice-versa. And for $\alpha < 0$, this effect is reversed.

Now, using the values of ρ_h and ρ_m from (13) and (17), the Eq. (16) can be simplified as

$$\dot{H} + \frac{3(\alpha+1)}{(3\alpha+2)}\left[H^2 + (2\alpha+1)c^2\omega_h H^{2-\Delta}\right] = 0. \tag{18}$$

Here, we have obtained a single evolution equation of H. For $\Delta = 0$ the BHDE get converted in HDE, thus we will obtain the same results as obtained for HDE (Singh and Kumar 2016). In the present model, we consider a maximal deformation from the standard Bekenstein-Hawking entropy which is given by $\Delta = 1$. On solving Eq. (18) for $\Delta = 1$, we obtain the value of H as give by

$$H = \frac{(2\alpha+1)c^2\omega_h}{c_0(2\alpha+1)c^2\omega_h e^{\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h t}{(3\alpha+2)}} - 1}, \tag{19}$$

where $\alpha \neq -\frac{2}{3}$ and c_0 is a constant of integration. Equation (19) can be rewritten as

$$H = ((2\alpha+1)c^2\omega_h H_0 e^{-\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_0)}{(3\alpha+2)}})$$

$$/((2\alpha+1)c^2\omega_h + H_0(1-e^{-\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_0)}{(3\alpha+2)}})), \quad (20)$$

where H_0 is the value of the Hubble parameter at the present cosmic time $t = t_0$, the time where the BHDE starts to dominate. Using $H = \frac{\dot{a}}{a}$ in Eq. (20), the cosmic scale factor a is obtained as

$$a = c_1\Big[(2\alpha+1)\omega_h c^2 + H_0\Big(1-e^{-\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_0)}{(3\alpha+2)}}\Big)\Big]^{\frac{(3\alpha+2)}{3(1+\alpha)}}, \quad (21)$$

where c_1 is an another constant of integration. One can rewrite Eq. (21) as follows

$$a = a_0\Big[1+\frac{H_0}{(2\alpha+1)\omega_h c^2}\times\Big(1-e^{-\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_0)}{(3\alpha+2)}}\Big)\Big]^{\frac{(3\alpha+2)}{3(1+\alpha)}} \quad (22)$$

where a_0 is the value of the scale factor at the present cosmic time $t = t_0$. Now, using Eq. (22) the value of H given by Eq. (20) can be written in terms of redshift z as

$$H(z) = \frac{H_0 - (2\alpha+1)\omega_h c^2[(1+z)^{-\frac{3(\alpha+1)}{(3\alpha+2)}} - 1]}{(1+z)^{-\frac{3(\alpha+1)}{(3\alpha+2)}}}. \quad (23)$$

After Hubble parameter H, the next important parameter is deceleration parameter (DP) q. The DP q, which is defined as $q = -\frac{\ddot{a}a}{\dot{a}^2}$, is a geometric parameter which describes the acceleration or deceleration of the universe depending on its negative or positive value, respectively. We have obtained the value of q as

$$q = \frac{3(1+\alpha)}{(3\alpha+2)}\Big[-1+\frac{1}{3(1+\alpha)} + \Big(\frac{(2\alpha+1)\omega_h c^2}{H_0}+1\Big)e^{\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_o)}{(3\alpha+2)}}\Big], \quad (24)$$

and in the term of redshift parameter z, the value of q can be written as

$$q(z) = \frac{3(1+\alpha)}{(3\alpha+2)}\Big[-1+\frac{1}{3(1+\alpha)} + \frac{(2\alpha+1)\omega_h c^2 + H_0}{H_0 - (2\alpha+1)\omega_h c^2((1+z)^{\frac{-3(1+\alpha)}{(3\alpha+2)}} - 1)}\Big]. \quad (25)$$

We have obtained a non-constant value of q which may explain the recent phase transition of the universe. It is evident from (24) that accelerated expansion may be obtained if $\frac{3(1+\alpha)(H_0^2+(2\alpha+1)c^2\omega_h)}{H_0(3\alpha+2)} < e^{\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t_0-t)}{(3\alpha+2)}}$. Moreover, we observe a constant value $q = -1$ throughout the evolution for coupling parameter $\alpha = -1$, i.e, a specific type of coupling between matter and geometry can accelerate the universe throughout the evolution.

We can define the effective EoS parameter using the relation

$$\omega_{\text{eff}} = -1 - \frac{1}{3}\frac{2a}{h}\frac{dh}{da}, \quad (26)$$

where $h = \frac{H}{H_0}$ is the weighted Hubble parameter. Using (23) into (26), we obtain

$$\omega_{\text{eff}} = -\Big(H_0\alpha(1+z)^{\frac{3(1+\alpha)}{2+3\alpha}} + c^2\omega_h(1+2\alpha)\Big(-2+\alpha\Big(-3+(1+z)^{\frac{3(1+\alpha)}{2+3\alpha}}\Big)\Big)\Big) / \Big((2+3\alpha)\Big(H_0(1+z)^{\frac{3(1+\alpha)}{2+3\alpha}} + c^2\omega_h(1+2\alpha)\Big(-1+(1+z)^{\frac{3(1+\alpha)}{2+3\alpha}}\Big)\Big)\Big). \quad (27)$$

From the above equation, it can be observed that as $z \to -1$ in the late-time ω_{eff} approaches to -1. It is to be noted that the ω_{eff} represents EoS contributed by all components of the universe. Thus, if $\omega_{\text{eff}} \to -1$ then it does not mean that the model is converging to Λ-CDM model.

Observational cosmology has emerged as a game changer in the field of cosmology. To check the viability of a cosmological model, it's observational confrontation is required. Therefore, we constrain the model parameters using MCMC method in next section where we use the observational data obtained from various observational probes. We will use observational constraining of these model parameters to discuss the evolution of the universe within our model in details.

4 Parameters estimation

In the domain of observational cosmology, a critical component involves the development of ideal cosmological models. Achieving this necessitates constraining the parameters of the model via thorough analysis of observational data. In this paper, the data set of choice is combination of measurements obtained with the latest pantheon+ data sample derived from type Ia supernovae observations, observational Hubble data and local Hubble constant from SH0ES.

4.1 Type Ia supernovae (Pantheon+ data)

A recent compilation of spectroscopically confirmed type Ia supernovae (SNIa), referred to as Pantheon+, has emerged. This dataset directly builds upon the Pantheon analysis

Table 1 Best-fit values of the model parameters

	c	ω_h	α	H_0	M
SNIa+HD	$-9.008^{+0.465}_{-0.671}$	$-0.873^{+0.115}_{-0.078}$	$-0.181^{+0.037}_{-0.032}$	$67.764^{+1.274}_{-1.354}$	$23.831^{+0.009}_{-0.007}$
SNIa+HD+SH0ES	$-9.474^{+0.809}_{-0.349}$	$-0.866^{+0.156}_{-0.130}$	$-0.170^{+0.069}_{-0.044}$	$70.440^{+0.816}_{-0.869}$	$23.823^{+0.009}_{-0.008}$
SNIa+HD+SH0ES (GR)	$-8.587^{+1.184}_{-0.924}$	$-0.577^{+0.105}_{-0.178}$	–	$70.646^{+1.054}_{-1.152}$	$23.832^{+0.008}_{-0.006}$

(Scolnic et al. 2018) and its predecessor, the Joint Light-curve Analysis. In the realm of cosmology, the key disparity between the Pantheon+ study and the original Pantheon analysis lies in the incorporation of new datasets. The Pantheon+ dataset comprises 1701 SNIa samples, an increase from the 1048 samples in the original Pantheon analysis. These 1701 light curves, representing 1550 unique SNIa within the redshift range of $0.001 < z < 2.26$, constitute the Pantheon+ dataset (Brout et al. 2022; Scolnic et al. 2022). The efficacy of SNIa as standard candles becomes apparent when assessing their constraining power.

The χ^2 for SNIa is defined as

$$\chi^2_{SNIa} = m^T C^{-1} m \tag{28}$$

where $m = m_B - m_{th}$ with

$$m_{th} = 5\, log_{10} D_L + M, \tag{29}$$

and C is the covariance matrix of μ_{obs}. The luminosity distance d_L is defined by $d_L = (c/H_0) D_L$, where

$$D_L = (1 + z_{hel}) \int_0^{z_{cmb}} \frac{H_0 dz'}{H(z', \theta)} \tag{30}$$

where θ represents the set of model parameters, z_{hel} is the heliocentric redshift and z_{cmb} is the redshift of the CMB rest frame.

4.2 Observational Hubble data (HD)

The investigation into the cosmic expansion history has drawn considerable interest from researchers, primarily due to the model-independent nature of measuring the Hubble parameter $H(z)$. In this paper, we utilize 36 measurements of $H(z)$, where the initial 31 measurements are derived from the cosmic chronometric method (Moresco et al. 2012), three measurements are correlated with the Baryon Acoustic Oscillations (BAO) signal in galaxy distribution (Alam et al. 2017), and the final two measurements are determined from the BAO signal in Lyman-alpha forest distribution alone or cross-correlated with quasistellar objects (QSOs) (Font-Ribera et al. 2014; Delubac et al. 2015).

The χ^2 for 33 Hubble measurement is given by

$$\chi^2_{CC+Ly\alpha} = \sum_{i=1}^{33} \frac{[H_{obs}(z_i) - H_{th}(z_i)]^2}{\sigma_i^2}, \tag{31}$$

In this context, $H_{th}(z_i)$ and $H_{obs}(z_i)$ denote the theoretical and observed values, respectively, while σ_i^2 represents the standard deviation associated with each $H_{obs}(z_i)$.

The χ^2 for the 3 galaxy distribution measurements is given by

$$\chi^2_{gal} = A^T C^{-1} A \tag{32}$$

where C is the covariance matrix given by (Alam et al. 2017)

$$C = \begin{bmatrix} 3.65 & 1.78 & 0.93 \\ 1.78 & 3.65 & 2.20 \\ 0.93 & 2.20 & 4.45 \end{bmatrix}$$

and

$$A = \begin{bmatrix} H_{obs}(0.38) - H_{th}(0.38) \\ H_{obs}(0.51) - H_{th}(0.51) \\ H_{obs}(0.61) - H_{th}(0.61) \end{bmatrix}$$

So, the combined χ^2 for HD is

$$\chi^2_{HD} = \chi^2_{CC+Ly\alpha} + \chi^2_{gal} \tag{33}$$

4.3 Local Hubble constant

Moreover, we incorporate the recent measurement of the local Hubble constant, denoted by $H_0 = 73.5 \pm 1.4$ km s^{-1} Mpc^{-1}, provided by the SH0ES collaboration, as documented in (Reid et al. 2019). Throughout this paper, we use the unit km s^{-1} Mpc^{-1} of Hubble constant.

In order to explore the parameter space, we employ python EMCEE package (Foreman-Mackey et al. 2013) and the MCMC sample technique. Upon minimization of a total χ^2, it will produce the best fits of model parameters $\{c, \omega_h, \alpha, H_0\}$ and M, which is a free normalization parameter used in SNIa observation. Further, we have considered the combinations $SNIa + HD$ and $SNIa + HD + SH0ES$ separately to constrain the model parameters. We have also obtained the best fit values of the model parameters in the case when $\alpha = 0$, i.e., in the case of GR. We present the observational summary in Table 1 and the contour plots for the same are shown in Figs. 1–3 with $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level. It is important to notice from contour plots for BHDE model in $f(R,T)$ gravity shown in Figs. 1-2 that the case $\alpha = 0$ is excluded at 2σ confidence level, i.e., the recent observations used in the present

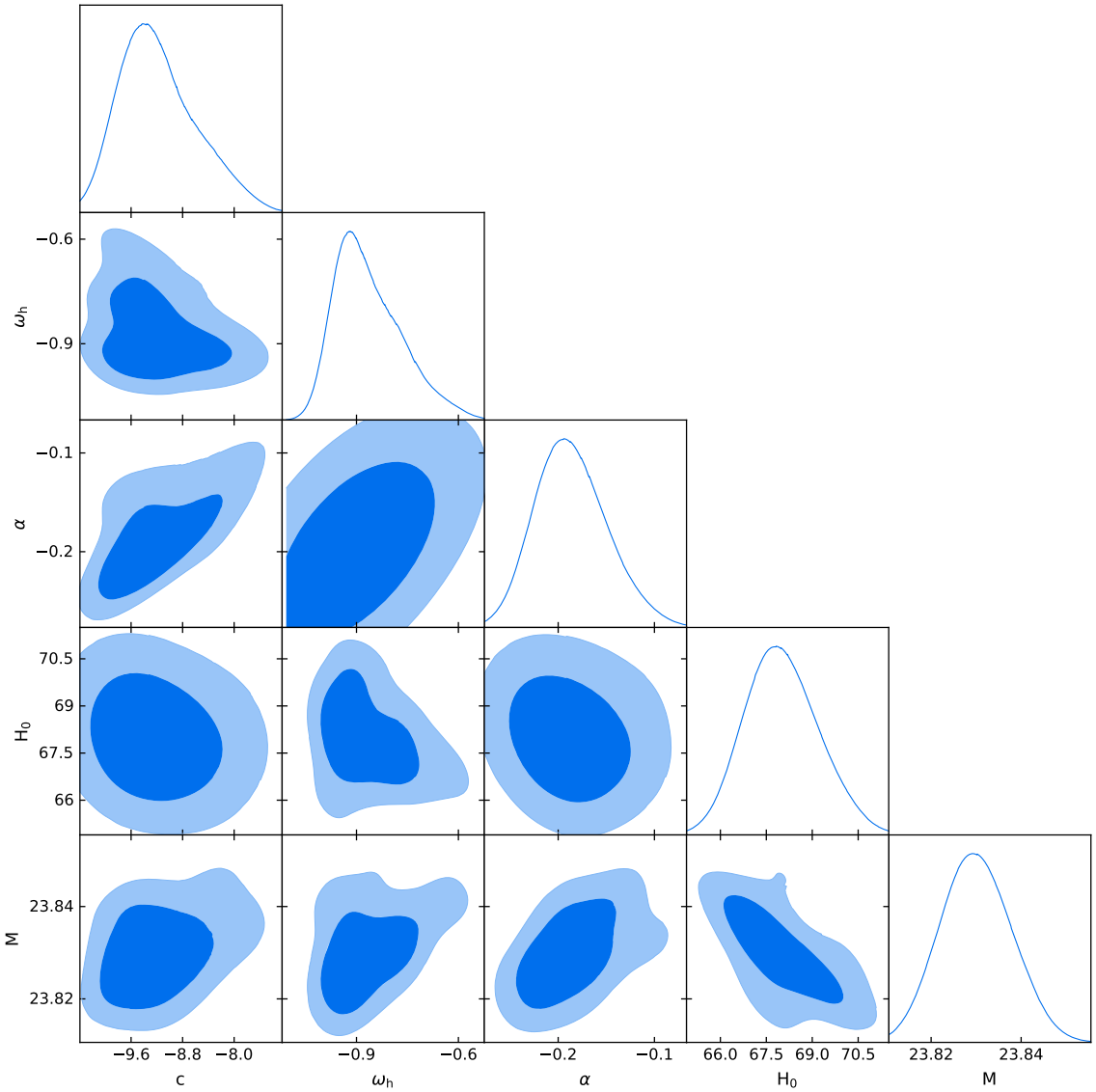


Fig. 1 1σ (68.3%) and 2σ (95.4%) confidence level contour plots for the model parameters using the observational data SNIa+HD

work are favoring $f(R,T)$ gravity over GR. It has also been shown in a recent paper (Bouali et al. 2023) that recent observations favor $f(R,T)$ gravity over GR.

5 Evolution of the model

In the Sect. 3, we have obtained the value of cosmological parameters like Hubble parameter H, DP q and effective EoS parameter ω_{eff} and discussed these parameters to some extent. Then, in Sect. 4, we applied two different combinations of recent observations to estimate the best fit values of the model parameter. Now, in the present section, we are interested to discuss the evolution of the universe through the obtained cosmological parameters in the light of obtained best fit values of the model parameters. We have divided the section into two subsections where in first subsection we discuss cosmological analysis through cosmological parameters and in second subsection we perform diagnostic analysis using statefinder and Om diagnostic.

5.1 Cosmological analysis

To study the cosmic evolution, there are many cosmological parameters which play an important role. Firstly, we plot the Hubble parameter along with the error bar and uncertainty region for the best fit values of the model parameters in Fig. 4. We compare our model with BHDE model in GR and Λ-CDM model. We find that the fitting achieved is compatible with the observational Hubble data at the low redshifts but their evolution with increasing z differs appreciably. For our BHDE model in $f(R,T)$ gravity, we have

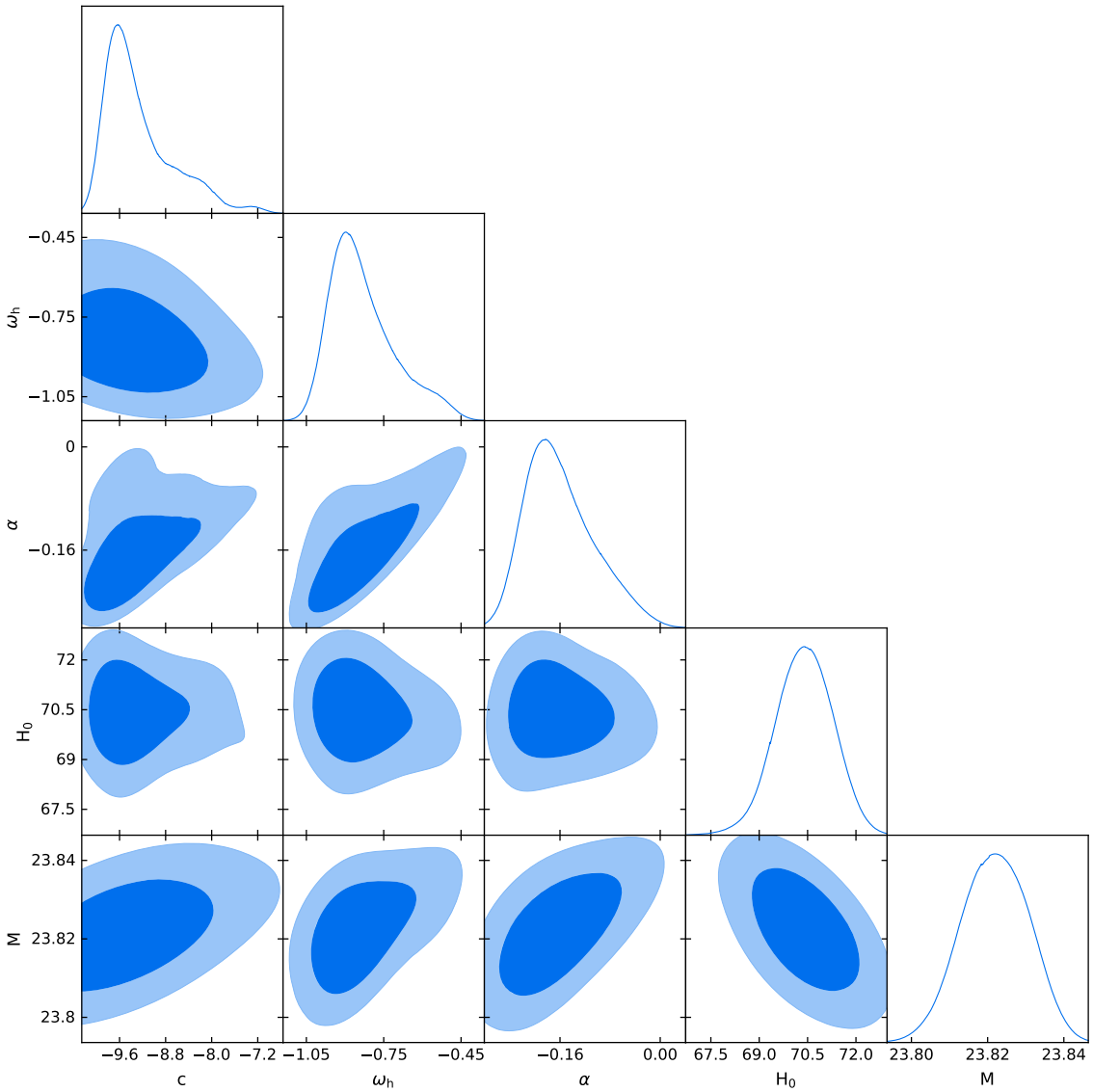


Fig. 2 1σ (68.3%) and 2σ (95.4%) confidence level contour plots for the model parameters using the observational data SNIa+HD+SH0ES

obtained the value of hubble constant $H_0 = 67.764^{+1.274}_{-1.354}$ when we consider combined data $SNIa + HD$ and we have obtained $H_0 = 70.440^{+0.816}_{-0.869}$ for $SNIa + HD + SH0ES$ data. We are observing a significant change in the value of H_0 when we add the data of $SH0ES$ collaboration. In the case of BHDE model in GR, we are getting almost same value $H_0 = 70.646^{+1.054}_{-1.152}$ for $SNIa + HD + SH0ES$ data but with large uncertainty region. The final data of Planck collaboration provided the value $H_0 = 67.4 \pm 0.5$ (Aghanim et al. 2020) taking Λ-CDM model as the model of evolution. In a very recent paper, authors (Riess 2022) have obtained $H_0 = 73.04 \pm 1.04$ using latest Cepheid+SNIa sample in a model independent analysis. We observe that the value of H_0 in our model for $SNIa + HD$ date is very close to the value from Λ-CDM model for Planck data. The value $H_0 = 70.440^{+0.816}_{-0.869}$ obtained from the data $SNIa + HD + SH0ES$ in our model is lying in between the values $H_0 = 67.4 \pm 0.5$ obtained from Λ-CDM model and the model independent value $H_0 = 73.04 \pm 1.04$. The significant difference between the values $H_0 = 67.4 \pm 0.5$ obtained from Λ-CDM model using Planck data and the model independent value $H_0 = 73.04 \pm 1.04$ is the reason of so called H_0 tension. Therefore, the BHDE model in $f(R,T)$ gravity could decrease the H_0 tension, bringing it to an acceptable range of error. It may be observed that the BHDE model in GR also able to decrease the H_0 tension. It has also been shown that HDE may explain the H_0 tension and authors obtained $H_0 = 71.54 \pm 1.78$ using Planck $CMB + BAO$ data (Dai et al. 2020). A detailed review of H_0 tension with the efforts that have been made to resolve it can be seen in (Valentino et al. 2021a,b).

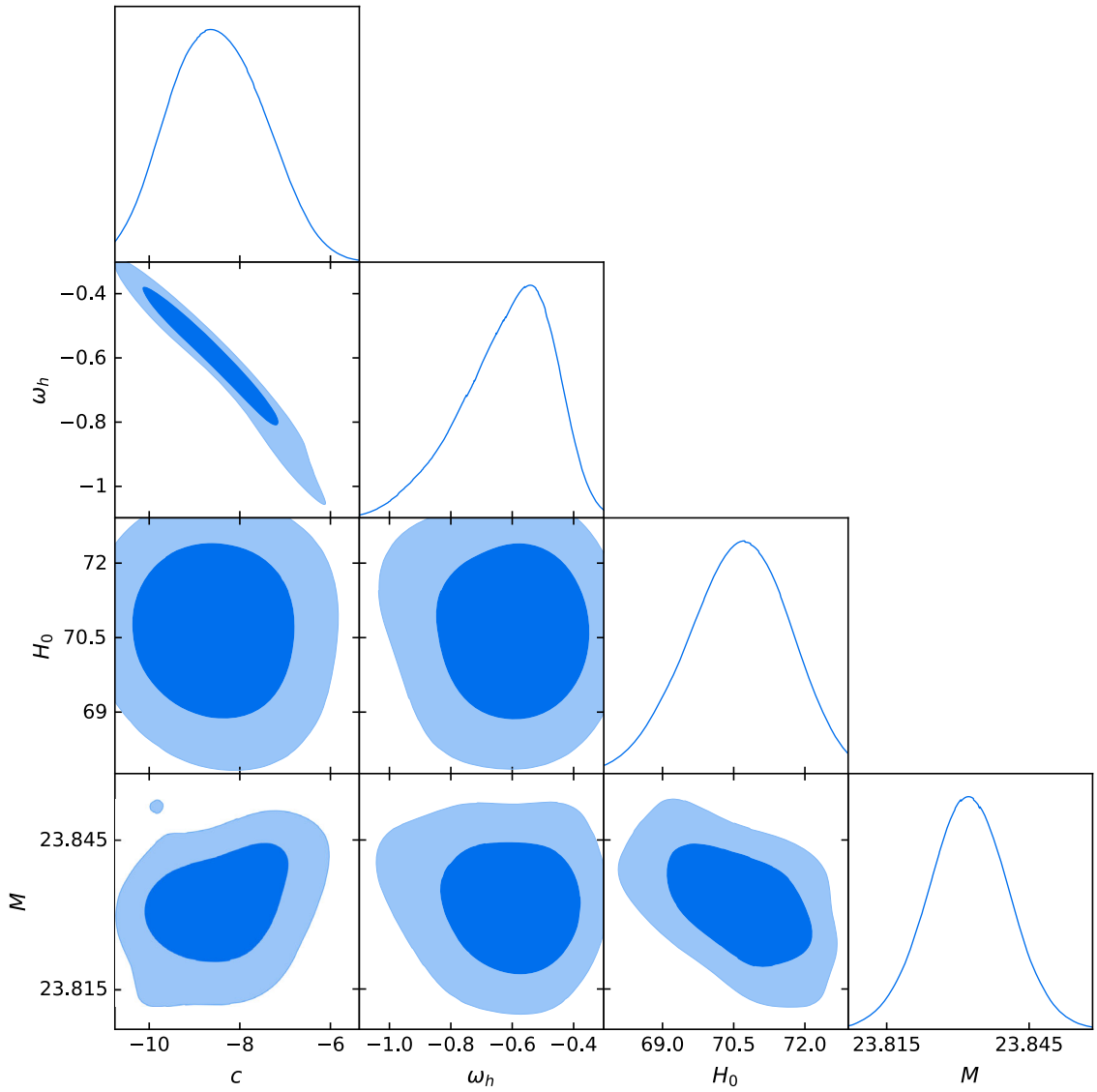


Fig. 3 1σ (68.3%) and 2σ (95.4%) confidence level contour plots for the model parameters using the observational data SNIa+HD+SH0ES for BHDE model in GR

We have plotted the graphs of $q(z)$ for obtained best fit values of the model parameters shown in Fig. 5. We observe a smooth phase transition of the universe from decelerated expansion to accelerated expansion which is compatible with recent observations. The phase transition occurs at transition redshift $z_t = 0.9$ for best fit value of the model parameters obtained from $SNIa + HD$ data and at $z_t = 1.3$ for $SNIa + HD + SH0ES$ data in our model. The transition redshift for the Λ-CDM model has been obtained as $z_t = 0.761 \pm 0.055$ (Lu et al. 2011) and we get $z_t = 1.1$ for $SNIa + HD + SH0ES$ data for BHDE model in GR. The present observational value of DP is given by $q_0 = -0.51 \pm 0.024$ (Riess 2022). One can observe from the Fig. 5 that the values of q_0 in our model is very close to the observational value.

Further, we have plotted effective EoS parameter for best fit values of the model parameters in Fig. 6 which shows that $\omega_{\text{eff}} \to -1$ in the far future ($z \to -1$) for BHDE model in $f(R, T)$ gravity and GR. It is also observed from the figure that our model is free from big-rip singularity as the curve does not cross phantom divide line. It is also important to note that ω_{eff} in not changing the sign in the case of BHDE model in GR as it is changing for $f(R, T)$ gravity, see Fig. 6. If we discuss about the EoS parameter of BHDE ω_h, the current value of ω_h is found to be $-0.873^{+0.115}_{-0.078}$ for $SNIa + HD$ data and $-0.866^{+0.156}_{-0.130}$ for $SNIa + HD + SH0ES$ data, which are in agreement with the value -0.93 predicted by the joint analysis of $WMAP + BAO + H_0 + SNIa$ data (Komatsu et al. 2011), while the value obtained from BHDE model in GR (taking $\alpha = 0$) using $SNIa + HD + SH0ES$

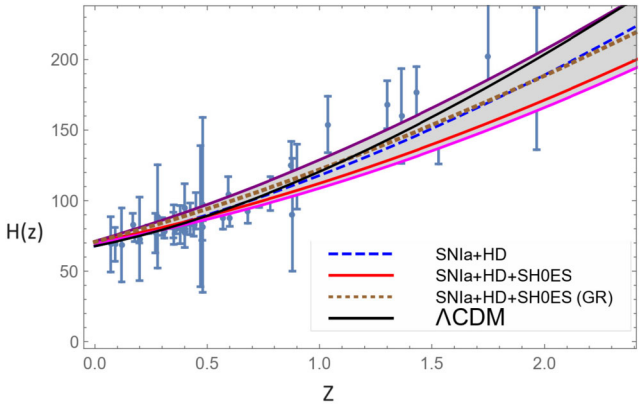


Fig. 4 The evolution of the Hubble parameter as a function of redshift z is depicted for the best-fit values of the BHDE model in both $f(R,T)$ gravity and GR for various data combinations. The uncertainty region for the SNIa+HD+SH0ES dataset is included at 1σ (68.3%) confidence level. The 36 observational $H(z)$ points are shown with grey error bars. Additionally, the variation of the Hubble function in the standard Λ-CDM model is represented by the black solid curve

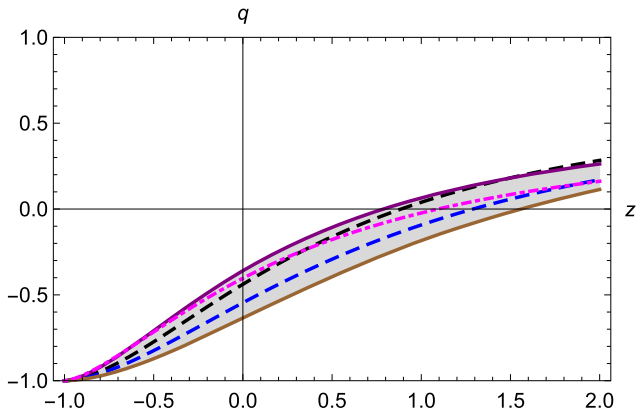


Fig. 5 We have plotted deceleration parameter against redshift z for best fit values of the model parameters where black dashed line corresponds to SNIa+HD data, blue dashed line correspond to SNIa+HD+SH0ES data and magenta dashed line represents trajectory for BHDE in GR for SNIa+HD+SH0ES data. The uncertainty region for the SNIa+HD+SH0ES dataset is included at 1σ (68.3%) confidence level

data is $-0.577^{+0.178}_{-0.105}$ which is very low in comparison of the observed value. Thus, BHDE model behaving significantly different in GR and $f(R,T)$ gravity on the basis of EoS parameters.

Next, we can calculate the age of the universe at the redshift z, which is given by $t(z) = T(z)/H_0$, where

$$T(z) = \int_z^\infty \frac{dz'}{(1+z')(H(z')/H_0)} \tag{34}$$

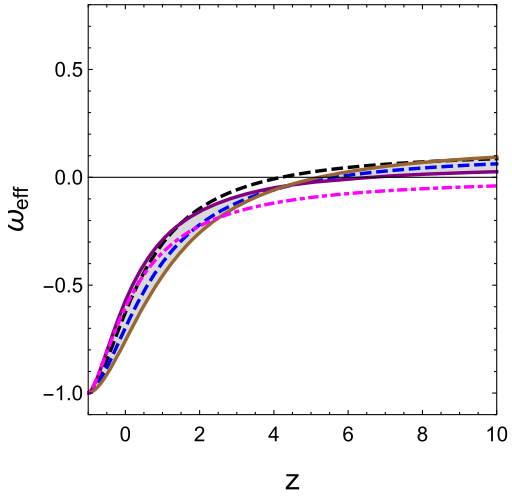


Fig. 6 The behavior of effective equation of state parameter as a function of z for best fit values of the model parameters has been plotted. The black dashed line corresponds to SNIa+HD data, blue dashed line correspond to SNIa+HD+SH0ES data and magenta dashed line represents trajectory for BHDE in GR for SNIa+HD+SH0ES data. The uncertainty region for the SNIa+HD+SH0ES dataset is included at 1σ (68.3%) confidence level

is dimensionless age parameter. For Λ-CDM model, the age parameter is (Feng and Li 2009)

$$T(z) = \int_z^\infty \frac{dz'}{(1+z')[\Omega_{m0}(1+z')^3 + (1-\Omega_{m0})]^{1/2}}. \tag{35}$$

With the help of above equations, we have plotted the age of the universe against the redshift z in Fig. 7. We have also plotted the age of the universe corresponding to concordance model (Λ-CDM model) for comparison with our model. The current age of the universe is found to be $t_0 = 14.09$ Gyr corresponding to $SNIa + HD$ data whereas it is found to be $t_0 = 15.43$ Gyr corresponding to $SNIa + HD + SH0ES$ data. It is observed that the age corresponding to the $SNIa + HD$ data is slightly higher but the age corresponding to $SNIa + HD + SH0ES$ data is comparatively much higher than the concordance model ($t_0 \sim 13.8$ Gyr). The age for BHDE model in GR is found to be $t_0 = 14.72$ Gyr corresponding to $SNIa + HD + SH0ES$ data which is also higher than the concordance model. The Λ-CDM model faces the age problem because stars have been found in the universe having age more than the age of universe predicted by the model. Moreover, the existence of early galaxies is also not compatible with the age of universe provided by the Λ-CDM model (see Gupta 2023 and references therein). The age of universe for the BHDE model in $f(R,T)$ gravity is higher than in GR corresponding to $SNIa + HD + SH0ES$ data. However, both models are able to resolve the age problem of Λ-CDM model but the BHDE model in $f(R,T)$ gravity seems more suitable as it provides higher value $t_0 = 15.43$ Gyr of the age of universe.

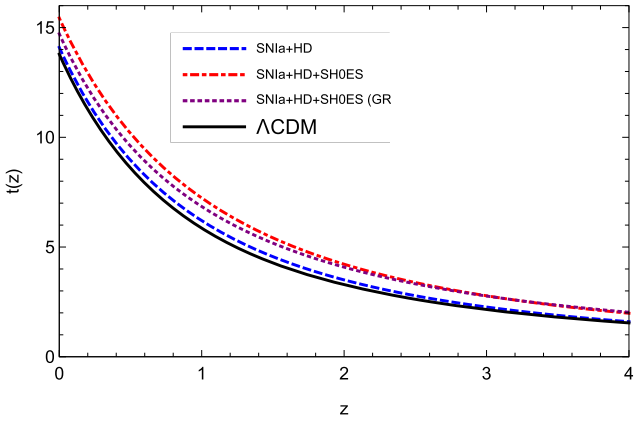


Fig. 7 The age of universe as a function of z plotted for best fit values of the model parameters. The blue dashed line corresponds to SNIa+HD data, red dot dashed line correspond to SNIa+HD+SH0ES data, violate dashed line represents trajectory for BHDE in GR for SNIa+HD+SH0ES data and solid black line corresponds to Λ-CDM model

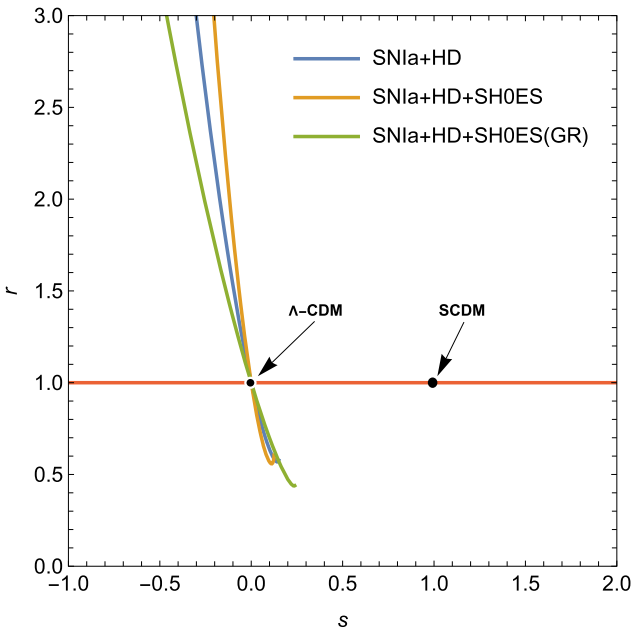


Fig. 8 The trajectories in $r - s$ plane are plotted for best fit values of the model parameters. The blue line corresponds to SNIa+HD data, magenta line corresponds to SNIa+HD+SH0ES data and green line represents trajectory for BHDE in GR for SNIa+HD+SH0ES data

Moreover, the observations also favouring $f(R,T)$ gravity over GR which has been discussed in Sect. 4.

5.2 Diagnostic analysis

Sahni et al. (2003) proposed a new diagnostic pair {r, s} to discriminate among DE models and the analysis is called the statefinder analysis. Alam et al. (2003) presented a detailed study on statefinder analysis. The statefinder is a geometrical diagnostic and allows us to characterize the properties of DE in a model-independent manner. The standard Λ-CDM model has the value of the statefinder pair {1, 0} and SCDM model has the value {1, 1}. Based on the behavior of trajectories in $r - s$ plane for a model, we can discriminate the model with other DE models. The statefinder pair {r, s} is defined as follows (Sahni et al. 2003)

$$r = \frac{\dddot{a}}{aH^3}, \quad s = \frac{(r-1)}{3(q-\frac{1}{2})}, \tag{36}$$

where the notations have their usual meaning. We have obtained the values of the statefinder parameters r and s for our model which are given by

$$r = \frac{e^{\frac{-6(1+\alpha)(2\alpha+1)c^2\omega_h t_0}{(3\alpha+2)}}}{H_0^2(2+3\alpha)^2}\Big[e^{\frac{6(1+\alpha)(2\alpha+1)c^2\omega_h t_0}{(3\alpha+2)}} H_0^2(2+3\alpha)^2$$
$$- 9e^{\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h (t+t_0)}{(3\alpha+2)}}$$
$$\times H_0(1+3\alpha+2\alpha^2)(H_0+c^2\omega_h(1+2\alpha))$$
$$+ 9e^{\frac{6(1+\alpha)(2\alpha+1)c^2\omega_h t}{(3\alpha+2)}}(1+\alpha)^2(H_0+c^2\omega_h(1+2\alpha))^2\Big],$$

$$s = 1/\Big(3H_0^2(2+3\alpha)^2$$
$$\times \Big(-\frac{1}{2}$$
$$+ \frac{e^{\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_0)}{(3\alpha+2)}}(H_0+3c^2\omega_h(1+3\alpha+2\alpha)^2)}{H_0(2+3\alpha)}\Big)\Big)$$
$$\Big[-9e^{\frac{3(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_0)}{(3\alpha+2)}}$$
$$\times H_0(1+3\alpha+2\alpha^2)(H_0+c^2\omega_h(1+2\alpha))$$
$$+9e^{\frac{6(1+\alpha)(2\alpha+1)c^2\omega_h(t-t_0)}{(3\alpha+2)}}(1+\alpha)^2(H_0+c^2\omega_h(1+2\alpha))^2\Big].$$

We have plotted trajectories in $r - s$ plane for our model shown in Fig. 8. The trajectories show Chaplygin gas type of behavior in starting, pass through Λ-CDM fixed point (1,0) during the evolution and then show quintessence type of behavior in the late evolution. Two trajectories in the figure have been plotted for best fit values of the model parameters obtained from $SNIa + HD$ data and $SNIa + HD + SH0ES$ data, where $t_0 = 1$ has been considered to plot the trajectories. We have also plotted trajectory of BHDE model in GR for best fit values of the model obtained from $SNIa + HD + SH0ES$ data. The trajectory shows qualitatively same behavior as in $f(R,T)$ gravity. We observe that the trajectories show small deviation in Chaplygin gas region but overall behavior of the trajectories remain same.

We have used another diagnostic parameter, called Om, which was introduced by Sahni et al. (2008) to distinguish various DE models from the Λ-CDM model. The slope of

$Om(z)$ can differentiate between different models of DE even if the value of the matter density is not accurately known. The Om diagnostic depends explicitly on the Hubble parameter H and is defined by

$$Om(z) = \frac{\frac{H^2(z)}{H_0^2} - 1}{(1+z)^3 - 1}. \tag{37}$$

With the help of (23), $Om(z)$ for our model can be obtained as

$$Om(z) = \frac{1}{(1+z)^3 - 1}\left[-1 + \frac{1}{H_0^2}\left\{ H_0(1+z)^{\frac{3(1+\alpha)}{2+3\alpha}} + c^2\omega_h(1+2\alpha)\left(-1 + (1+z)^{\frac{3(1+\alpha)}{2+3\alpha}} \right) \right\}^2 \right]. \tag{38}$$

To analyse the behavior of $Om(z)$, we plot the graph in $Om(z) - z$ plane for the best fit values of model parameters in Fig. 9. It is to be noted that the positive slope of $Om(z)$ represents phantom like behavior of DE whereas the negative slope represents the quintessence behavior. It can be observed from the figure that the trajectories of $Om(z)$ have the negative slope which represents the behavior similar to the quintessence model. The trajectory of the BHDE model in GR also shows similar behavior but it has more negative slope.

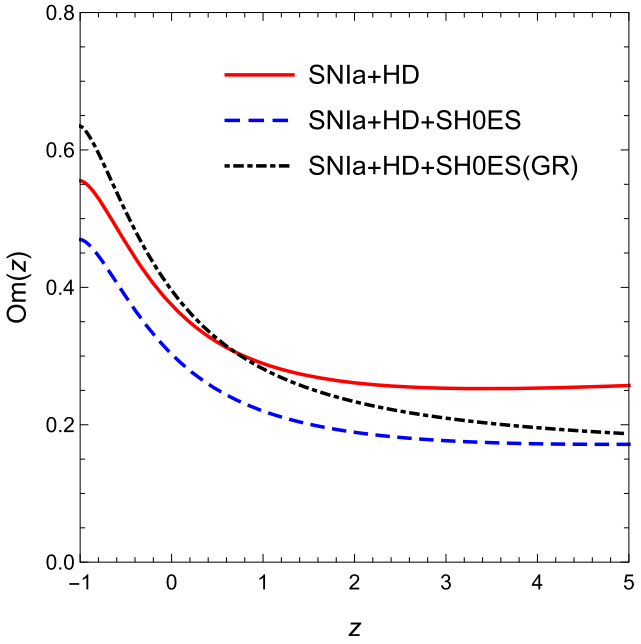


Fig. 9 The trajectories in $Om(z) - z$ plane are plotted for best fit values of the model parameters. The solid red line corresponds to SNIa+HD data, blue dashed line correspond to SNIa+HD+SH0ES data and black dot dashed line represents trajectory for BHDE in GR for SNIa+HD+SH0ES data

6 Conclusion

In the present work, we have analyzed a modified version of HDE based on Barrow entropy known as BHDE in the framework of $f(R,T)$ theory. We have considered a flat FLRW universe to discuss the model which is filled with BHDE and DM. The Hubble horizon has been considered as IR cut-off to obtain the energy density of BHDE because it is simplest and causal free choice for IR cut-off. We obtained the field equation of BHDE model in the framework of $f(R,T)$ theory and observed that the field equations can be solved analytically for only two values of BHDE parameter $\Delta = 0, 1$. For $\Delta = 0$, BHDE reduces to HDE and one will get the same result as obtained in (Singh and Kumar 2016). We consider a maximal deformation from the standard Bekenstein-Hawking entropy which is given by $\Delta = 1$ to obtain analytical solution of the field equations. In process to analyse the cosmic evolution, we have obtained the values of Hubble parameter and cosmic scale factor. Further, we have obtained the DP q and effective EoS parameter ω_{eff} using the values of Hubble parameter and cosmic scale factor.

In the present scenario, the observations are playing a crucial role in cosmology. Enormous amount of data being received every year and new probes being launched to explore every aspects of the universe. Therefore, it becomes utmost important to explore a theoretical model of cosmic evolution using available observational data. We have explored the model observationally and obtained best fit values of the model parameters using the MCMC method. The BHDE model in the framework of GR has been discussed (Adhikary et al. 2021; Oliveros et al. 2022) and has been constrained using observational data (Anagnostopoulos et al. 2020). Our BHDE model in $f(R,T)$ gravity gets converted in the BHDE model in GR for the parameter value $\alpha = 0$. Therefore, we have also constrained the BHDE model in GR taking $\alpha = 0$ in our model using the latest observational data to compare the results of both models. We have used two combinations $SNIa + HD$ and $SNIa + HD + SH0ES$ of the latest observational data to constrain model parameters. We observed that the best fit values of H_0 and ω_h with in BHDE model in $f(R,T)$ gravity are in good agreement with latest observational values of the parameters. Contour plots for the same are shown in Figs. 1–3 with $1\sigma(68.3\%)$ and $2\sigma(95.4\%)$ confidence level. The BHDE model in $f(R,T)$ gravity is able to reduce the Hubble tension. It is important to note that the case $\alpha = 0$, where $f(R,T)$ gravity gets converted in GR, is excluded at 2σ confidence level, i.e., the recent observations are favoring $f(R,T)$ gravity over GR.

We have plotted graphs of q against redshift parameter z for observationally estimated values of model parameters. The graphs show a smooth phase transition from early decelerated expansion to recent accelerated expansion. We have observed that the phase transition occurs at $z_t = 0.9$ for best fit value of the model parameters obtained from $SNIa + HD$ data and at $z_t = 1.3$ for $SNIa + HD + SH0ES$ data which shows compatibility with the observations. The current value of ω_h is found to be $-0.873^{+0.115}_{-0.078}$ for $SNIa + HD$ data and $-0.866^{+0.156}_{-0.130}$ for $SNIa + HD + SH0ES$ data, which are in agreement with the value -0.93 predicted by the joint analysis of $WMAP + BAO + H_0 + SNIa$ data (Komatsu et al. 2011). While, the value $-0.577^{+0.105}_{-0.178}$ obtained for BHDE model in GR for $SNIa + HD + SH0ES$ data is not compatible with the above values.

The current age of the universe is found to be $t_0 = 14.09\ Gyr$ corresponding to $SNIa + HD$ data set whereas it is found to be $t_0 = 15.43\ Gyr$ corresponding to $SNIa + HD + SH0ES$ data set. It is found to be $t_0 = 14.72\ Gyr$ corresponding to $SNIa + HD + SH0ES$ data set for BHDE model in GR. The age of the universe for BHDE model within $f(R, T)$ gravity is higher than in GR corresponding to $SNIa + HD + SH0ES$ data. However, both the model are able to resolve the age problem of Λ-CDM model but the BHDE model in $f(R, T)$ gravity seems more suitable. Moreover, we have analyzed our model using statefinder analysis and Om diagnostic to compare our model with other existing DE models. The statefinder analysis shows that the model shows similarity with Chaplygin gas in early phase evolution, go through the fixed point value of Λ-CDM model and shows quintessence type of behavior in late time evolution. It is to be noted that statefinder analysis of HDE model in $f(R, T)$ gravity has showed that the model behaves like quintessence model throughout the evolution (Singh and Kumar 2016). Thus, BHDE presents more richer phenomenon than HDE as discussed by Saridakis (2020). The BHDE model in GR presents qualitatively similar behavior as in $f(R, T)$ gravity as far as the statefinder and Om diagnostics are concerned.

Acknowledgements We are grateful to Prof. A. Beesham for fruitful discussion and help to improve the quality of the manuscript. We are thankful to the anonymous referees for their helpful comments and suggestions, which have enhanced the completeness of this manuscript.

Author contributions K.D. wrote the main manuscript text and all other authors contributed to the final version of it. A.K. complied with the observational data analysis. The analysis of the manuscript was completed by all authors. The graphs was made by all the authors.

Funding No funding was obtained to complete this research work.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Ade, P.A.R., et al.: A & A **571**, A16 (2014)
Adhikary, P., et al.: Phys. Rev. D **104**, 123519 (2021)
Aghanim, N., et al.: A & A **641**, A6 (2020)
Aktas, C.: Mod. Phys. Lett. A **34**, 1950098 (2019)
Alam, U., et al.: Mon. Not. R. Astron. Soc. **344**, 1057 (2003)
Alam, S., et al.: Mon. Not. R. Astron. Soc. **470**, 2617 (2017)
Anagnostopoulos, F.K., Basilakos, S., Saridakis, E.N.: Eur. Phys. J. C **80**, 86 (2020)
Bahamonde, S., Bohmer, C.G., Wright, M.: Phys. Rev. D **92**, 104042 (2015)
Bamba, K., Nojiri, S., Odintsov, S.D.: J. Cosmol. Astropart. Phys. **0180**, 045 (2008)
Bamba, K., et al.: Eur. Phys. J. C **67**, 295 (2010)
Bertini, N.R., Velten, H.: Phys. Rev. D **107**, 124005 (2023)
Bhattacharjee, S., et al.: Eur. Phys. J. Plus **135**, 576 (2020)
Bouali, A., et al.: Mon. Not. R. Astron. Soc. **526**, 4192 (2023)
Brout, D., et al.: Astrophys. J. **938**(2), 110 (2022)
Buchdahl, H.A.: Mon. Not. R. Astron. Soc. **150**, 1 (1970)
Campo, S., et al.: Phys. Rev. D **83**, 123006 (2011)
Carroll, S.M.: Living Rev. Relativ. **4**, 1 (2001)
Clifton, T., et al.: Phys. Rep. **513**, 1 (2012)
Copeland, E.J., Sami, M., Tsujikawa, S.: Int. J. Mod. Phys. D **15**, 1753 (2006)
Dai, W.M., Ma, Y.Z., He, H.J.: Phys. Rev. D **102**, 121302(R) (2020)
Delubac, T., et al.: Astron. Astrophys. **574**, A59 (2015)
Elizalde, E., et al.: Class. Quantum Gravity **27**, 095007 (2010)
Feng, C.J., Li, X.Z.: Phys. Lett. B **680**, 355 (2009)
Font-Ribera, A., et al.: J. Cosmol. Astropart. Phys. **05**, 027 (2014)
Foreman-Mackey, D., et al.: Publ. Astron. Soc. Pac. **125**, 306 (2013)
Gupta, R.P.: Mon. Not. R. Astron. Soc. **524**, 3385 (2023)
Harko, T.: Phys. Rev. D **90**, 044067 (2014)
Harko, T., et al.: Phys. Rev. D **84**, 024020 (2011)
Houndjo, M.J.S.: Int. J. Mod. Phys. D **21**, 1250003 (2012)
Hsu, S.D.H.: Phys. Lett. B **594**, 13 (2004)
Komatsu, E., et al.: Astrophys. J. Suppl. **192**, 18 (2011)
Kumar, P., Singh, C.P.: Astrophys. Space Sci. **357**, 120 (2015)
Li, M.: Phys. Lett. B **603**, 1 (2004)
Lu, J., Xu, L., Liu, M.: Phys. Lett. B **699**, 246 (2011)
Mamon, A.A., Paliathanasis, A., Saha, S.: Eur. Phys. J. Plus **136**, 134 (2021)
Matos, I.S., Calvao, M.O., Waga, I.: Phys. Rev. D **103**, 104059 (2021)
Moresco, M., et al.: J. Cosmol. Astropart. Phys. **08**, 006 (2012)
Naser, A., Pradhan, A.: Int. J. Theor. Phys. **53**, 289 (2014)
Nojiri, S., Odintsov, S.D.: Phys. Rep. **505**, 59 (2011)
Odintsov, S.D., Oikonomou, V.K., Banerjee, N.: Nucl. Phys. B **938**, 935 (2019)
Oliveros, A., Sabogal, M.A., Acero, M.A.: Eur. Phys. J. Plus **137**, 783 (2022)
Paul, B.C., Roy, B.C., Saha, A.: Eur. Phys. J. C **82**, 76 (2022)
Pavón, D., Zimdahl, W.: Phys. Lett. B **628**, 206 (2005)
Peebles, P.J.E., Ratra, B.: Rev. Mod. Phys. **75**, 559 (2003)
Perlmutter, S., et al.: Astrophys. J. **517**, 565 (1999)
Pinki, Kumar, P.: Gen. Relativ. Gravit. **55**, 46 (2023)
Pinto, M.A.S., Harko, T., Lobo, F.S.N.: Phys. Rev. D **106**, 044043 (2022)
Rajabia, F., Nozarib, K.: Eur. Phys. J. C **81**, 247 (2021)
Reid, M.J., Pesce, D.W., Riess, A.G.: Astrophys. J. **886**(2), L27 (2019)
Riess, A.G.: Astrophys. J. Lett. **943**, L2 (2022)

Riess, A.G., et al.: Astron. J. **116**, 1009 (1998)
Rosa, J.L.: Phys. Rev. D **103**, 104069 (2021)
Rosa, J.L., Kull, P.M.: Eur. Phys. J. C **82**, 1154 (2022)
Sahni, V., et al.: J. Exp. Theor. Phys. Lett. **77**, 201 (2003)
Sahni, V., Shafieloo, A., Starobinsky, A.A.: Phys. Rev. D **78**, 103502 (2008)
Sahni, V., Shafieloo, A., Starobinsky, A.A.: Astrophys. J. Lett. **793**, L40 (2014)
Sahoo, P.K., Mishra, B., Tripathy, S.K.: Indian J. Phys. **90**, 485 (2016)
Samart, D., Silasan, B., Channuie, P.: Phys. Rev. D **104**, 063517 (2021)
Sardar, G., Bose, A., Chakraborty, S.: Eur. Phys. J. C **83**, 41 (2023)
Saridakis, E.N.: Phys. Rev. D **102**, 123525 (2020)
Scolnic, D.M., et al.: Astrophys. J. **859**(2), 101 (2018)
Scolnic, D.M., et al.: Astrophys. J. **938**(2), 113 (2022)
Sharif, M., Nawazish, I.: Eur. Phys. J. C **77**, 198 (2017)
Sheykhi, A.: Phys. Lett. B **681**, 205 (2009)
Singh, C.P., Kumar, P.: Eur. Phys. J. C **74**, 3070 (2014)
Singh, C.P., Kumar, P.: Astrophys. Space Sci. **361**, 157 (2016)
Singh, C.P., Kumar, P.: Int. J. Theor. Phys. **56**, 3297 (2017)
Singh, C.P., Kumar, A.: Mod. Phys. Lett. A **33**, 1850225 (2018)
Singh, C.P., Singh, V.: Gen. Relativ. Gravit. **46**, 1696 (2014)
Spergel, D.N., et al.: Astrophys. J. Suppl. Ser. **148**, 175 (2003)
Spergel, D.N., et al.: Astrophys. J. Suppl. Ser. **170**, 377 (2007)
Srivastava, S., Sharma, U.K.: Int. J. Geom. Methods Mod. Phys. **18**, 2150014 (2020)
't Hooft, G.: (1993). arXiv:gr-qc/9310026
Tegmark, M., et al.: Phys. Rev. D **69**, 103501 (2004a)
Tegmark, M., et al.: Astrophys. J. **606**, 702 (2004b)
Valentino, E.D., et al.: Class. Quantum Gravity **38**, 153001 (2021a)
Valentino, E.D., et al.: Astropart. Phys. **131**, 102605 (2021b)
Wright, M.: Phys. Rev. D **93**, 103002 (2016)
Xu, Y., et al.: Eur. Phys. J. C **79**, 708 (2019)
Zimdahl, W., Pavón, D.: Class. Quantum Gravity **24**, 5461 (2007)
Zubair, M., Durrani, L.R.: Chin. J. Phys. **69**, 153 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Base and Exponent Prediction in Mathematical Expressions using Multi-Output CNN

1st Md Laraib Salam
Department of Electrical Engineering
Delhi Technological University
New Delhi, India
laraibsalam2013@gmail.com

2nd Akash S Balsaraf
Department of Information Technology
School of Engineering and Technology DYPU
Pune, India
akashsbalsaraf@gmail.com

3rd Gaurav Gupta
Department of Aerospace Engineering
IIST
Thiruvananthapuram, India
gauravxpgupta@gmail.com

arXiv:2407.14967v1 [cs.CV] 20 Jul 2024

***Abstract*—The use of neural networks and deep learning techniques in image processing has significantly advanced the field, enabling highly accurate recognition results. However, achieving high recognition rates often necessitates complex network models, which can be challenging to train and require substantial computational resources. This research presents a simplified yet effective approach to predicting both the base and exponent from images of mathematical expressions using a multi-output Convolutional Neural Network (CNN). The model is trained on 10,900 synthetically generated images containing exponent expressions, incorporating random noise, font size variations, and blur intensity to simulate real-world conditions. The proposed CNN model demonstrates robust performance with efficient training time. The experimental results indicate that the model achieves high accuracy in predicting the base and exponent values, proving the efficacy of this approach in handling noisy and varied input images.**

***Index Terms*—convolution neural network; handwritten digit recognition; combinatorial network; Gabor filter**

I. Introduction

Optical character recognition (OCR) technology encompasses both handwritten and printed character recognition, with applications ranging from postal code and financial value identification to tax form recognition and e-commerce digital processing. While significant progress has been made in these areas, achieving perfect recognition accuracy remains an ongoing challenge due to the variability and complexity of input images. In recent years, deep learning techniques, particularly Convolutional Neural Networks (CNNs), have revolutionized the field of image processing. CNNs are characterized by their local connectivity, weight sharing, and pooling operations, which enable them to extract robust and descriptive features from images. These attributes make CNNs well-suited for tasks requiring high accuracy in image recognition. This research focuses on a specialized application of CNNs: predicting the base and exponent values from images of mathematical expressions. Traditional approaches to OCR often struggle with the unique challenges posed by these images, which can include random noise, varying font sizes, and blur intensity. To address these challenges, we propose a multi-output CNN model that can simultaneously predict both the base and exponent from a single image. Our model is trained on a synthetically generated dataset of 10,900 images containing exponent expressions, designed to simulate real-world conditions with added noise and variations. Data augmentation techniques are employed to further enhance the model's generalization capabilities. The results demonstrate that our approach achieves high accuracy, offering a practical solution for recognizing mathematical expressions in noisy and varied contexts. This paper outlines the design and implementation of the multi-output CNN, detailing the pre processing steps, model architecture, and training process. Experimental results validate the effectiveness of our approach, highlighting the potential for applying similar techniques to other complex image recognition tasks.

II. Principles of the Method

A. Multi-Output CNN for Exponential Expression Recognition

In this paper, the methodology is divided into two main parts: model training and model testing. For model training, the Convolutional Neural Network (CNN) continuously updates the network parameters based on a synthetic dataset of 10,900 images. The dataset includes various forms of noise, font size variations, and blur to enhance the model's robustness. For model testing, the system evaluates the performance of the trained network on unseen images, predicting the base and exponent values, and demonstrates the model's generalization capability.

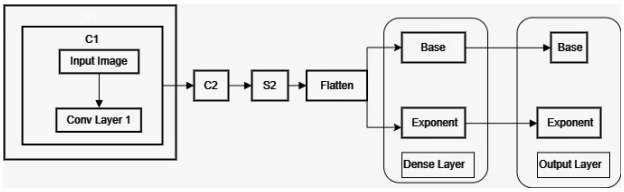


Fig. 1. CNN architecture for mathematical expression recognition.

B. Structure of the CNN

The CNN architecture is designed to process two-dimensional input matrices and is composed of several key layers:

- **Convolutional Layers**: The network begins with convolutional layers that extract hierarchical features from input images. Each convolutional layer applies filters to the input to produce feature maps.

1) Convolution Operation:

$$z_{i,j,k} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{i+m,j+n} \cdot w_{m,n,k} + b_k \quad (1)$$

where:

- $z_{i,j,k}$ is the output of the k-th filter at position (i, j),
- $x_{i+m,j+n}$ is the input pixel value at position $(i + m, j + n)$,
- $w_{m,n,k}$ is the weight of the k-th filter at position (m, n),
- b_k is the bias term for the k-th filter,
- $M \times N$ is the filter size.

- **Pooling Layers**: After convolution, pooling layers downsample the feature maps, reducing spatial dimensions while retaining important features. Max pooling is used in this architecture.
- **Fully Connected Layers**: Flattened outputs from the convolutional and pooling layers are passed through fully connected layers to generate the final predictions for base and exponent.

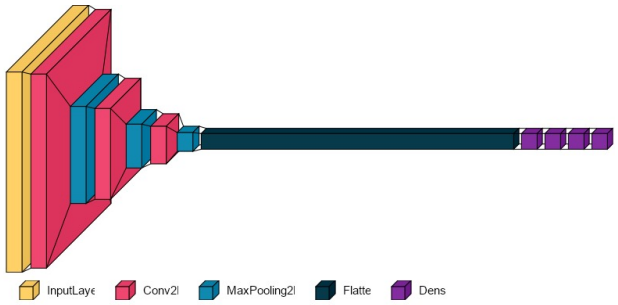


Fig. 2. Detailed CNN architecture of the model.

The convolution operation in a single layer is defined as:

$$(I * K)(i, j) = \sum_m \sum_n I(i + m, j + n)K(m, n) \quad (2)$$

where I is the input image and K is the convolution kernel. The Rectified Linear Unit (ReLU) activation function is used:

$$\text{ReLU}(x) = \max(0, x) \quad (3)$$

The CNN consists of multiple convolutional and pooling layers. The input pixel matrix is processed through convolution layers with various filters, followed by pooling operations to reduce dimensions. The final output is vectorized for classification. The convolution and pooling layers are trained to optimize the model's performance using algorithms like gradient descent.

C. Forward Propagation and Softmax Function

Forward propagation in the CNN involves passing the input through the network layers to obtain the output predictions. This process includes convolution, activation, pooling, and fully connected layers.

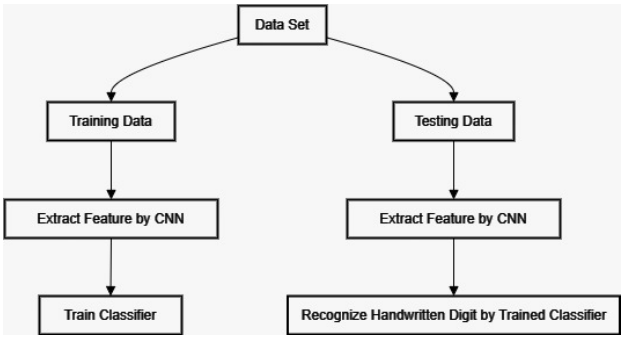


Fig. 3. Forward propagation through the CNN.

The softmax function is used to map the CNN's output into probability distributions:

$$\text{softmax}(V_i) = \frac{e^{V_i}}{\sum_j e^{V_j}} \quad (4)$$

where V_i represents the score for class i.

For multi-class classification, the loss function employed is sparse categorical cross-entropy:

$$\text{Loss} = -\sum_i y_i \log(\hat{y}_i) \quad (5)$$

where y_i is the true label and $\hat{y}_i$ is the predicted probability. The gradient descent method is used to optimize the model by updating the gradients with respect to the loss function.

III. Datasets

The dataset used for training and testing the CNN consists of 10,900 synthetic images of exponential expressions. These images are generated to include variations in noise, font size, and blur, designed to challenge and evaluate the model's performance. Each image is labeled with the base and exponent values to facilitate supervised learning.

IV. Experiments

Our experiments are divided into two main parts: training and testing.

A. Training

For the training phase, we utilized a convolutional neural network (CNN) model to learn from a dataset of 10,000 synthetically generated images. These images were created to simulate real-world conditions, incorporating variations in random noise, font size, and blur intensity. The CNN was trained using this dataset with the goal of predicting both the base and exponent values from images of mathematical expressions.

The network was trained over a total of 50 epochs. During training, the model's parameters were updated based on the loss function, which was computed using the cross-entropy loss for both the base and exponent outputs. The training process involved optimizing the model using the Adam optimizer and monitored with early stopping to prevent overfitting.

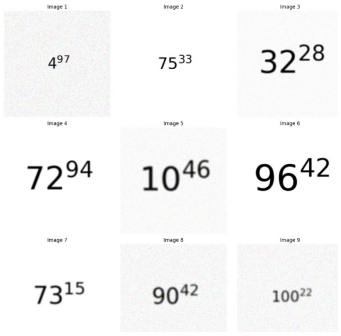


Fig. 4. Sample images from the test dataset with varying levels of noise and blur.

B. Testing

For the testing phase, we evaluated the performance of the trained CNN model using a separate test dataset. This dataset consisted of 1,000 images that were not seen by the model during the training phase. The test images were processed similarly to the training images, including normalization and resizing.

The performance of the model was assessed based on its accuracy in predicting the base and exponent values from the test images. Additionally, we evaluated the model's robustness by introducing varying levels of noise and blur to the test images to simulate real-world conditions.

Figure 4 shows some of the images from our test dataset, which include various levels of noise and blur to test the model's robustness.

We also analyzed the distribution of the base and exponent values in the test dataset to ensure a balanced representation. Figure 5 and Figure 6 illustrate these distributions.

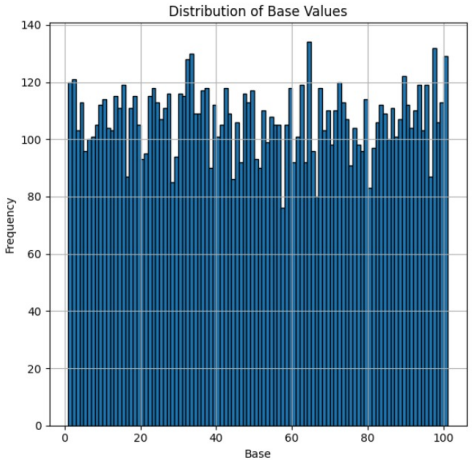


Fig. 5. Distribution of base values in the test dataset.

To further evaluate the model's performance, we examined the distribution of image attributes such as blur and noise levels in the test dataset. Figure 7 and Figure 8 show these distributions.

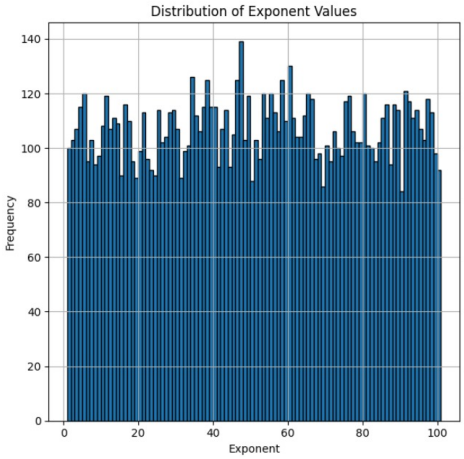


Fig. 6. Distribution of exponent values in the test dataset.

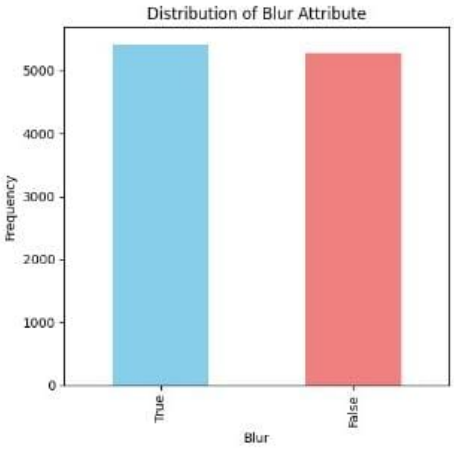


Fig. 7. Distribution of blur levels in the test dataset.

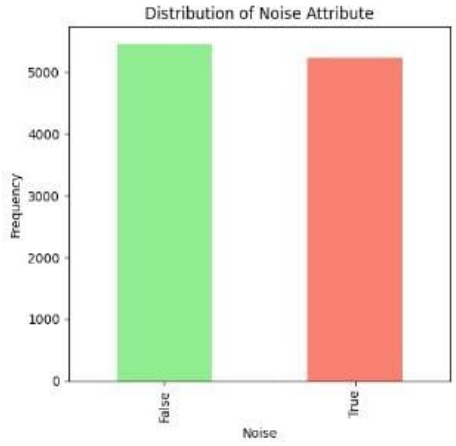


Fig. 8. Distribution of noise levels in the test dataset.

The model is able to accurately predict the base and exponent values from the images, demonstrating its robustness and effectiveness in handling real-world conditions. The results

from the testing phase confirm that the trained CNN model performs well even with variations in noise and blur levels, maintaining a high level of accuracy in its predictions.

The experimental results showed that the model achieved high accuracy in predicting both the base and exponent values. The effectiveness of the CNN was demonstrated through its ability to handle noisy and varied input images, maintaining a high level of performance across different test scenarios.

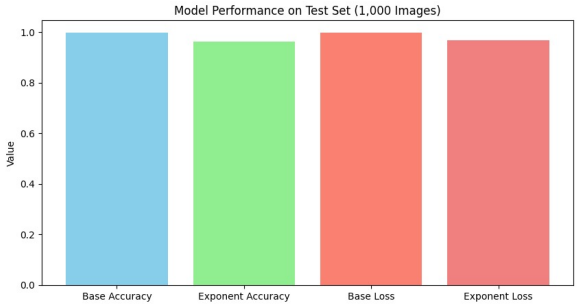


Fig. 9. Results of the testing phase showing model performance on various test scenarios.

V. Conclusion

In this paper, we have presented a comprehensive system for predicting base and exponent values from images of mathematical expressions using a multi-output Convolutional Neural Network (CNN). Our method demonstrates high accuracy and robustness in recognizing and predicting these values, even under varying conditions such as different levels of noise and blur.

The trained model was evaluated using a Python-generated dataset of 10,000 images, and testing showed that the model maintained high performance in predicting the desired outputs. The experimental results confirmed that our approach is effective and efficient, handling real-world variations and maintaining a high level of accuracy.

Compared to traditional methods, such as Histogram of Oriented Gradient (HOG), our CNN-based approach provides significant advantages in terms of accuracy, speed, and robustness. The ability of the model to generalize well to unseen data and to handle different image attributes makes it a strong candidate for practical applications in various fields.

In the future, we plan to extend our work by incorporating more diverse datasets to further improve the model's generalizability. Additionally, we aim to optimize the model architecture and explore advanced techniques such as transfer learning to enhance the performance and efficiency of the system. We also intend to integrate real-time processing capabilities to facilitate immediate feedback and predictions, making the system more applicable in dynamic and interactive environments.

Overall, our research demonstrates the potential of deep learning techniques in solving complex recognition tasks, and we are optimistic about the future advancements and applications of our work.

Acknowledgments

The authors would like to express their sincere gratitude to the anonymous reviewers and the editor for their insightful comments and suggestions, which have significantly contributed to enhancing the presentation and theoretical aspects of this paper. Their valuable feedback has been instrumental in improving the quality and clarity of this work.

References

[1] Scherer, D., Müller, A., Behnke, S. (2010). *Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition*. In International Conference on Artificial Neural Networks (ICANN), pages 92–101.

[2] Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems (NeurIPS), pages 1097–1105

[3] Zeiler, M. D., Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. In European Conference on Computer Vision (ECCV), pages 818–833.

[4] Y. Hou and H. Zhao, "Handwritten digit recognition based on depth neural network," in 2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), Okinawa, 2017 pp. 35-38.

[5] Yu, Wei, et al. "Exploiting the complementary strengths of multi-layer CNN features for image retrieval." Neurocomputing 237 (2017): 235-241.

[6] [1]C. Zhang, Z. Zhou, and L. Lin, "Handwritten Digit Recognition Based on Convolutional Neural Network", IEEE Xplore, Nov. 01, 2020.

2024 IEEE Intelligent Vehicles Symposium (IV)
June 2-5, 2024. Jeju Island, Korea

Camera-based Mobility Framework for Visually Impaired Pedestrians in Unstructured Environments

Aditya Kumar[1*], Anwesha Chakravarty[1*], Ayesha Choudhary[1], S. Indu[2]

2024 IEEE Intelligent Vehicle Symposium (IV) | 979-8-3503-4881-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/IV55156.2024.10588516

Abstract— **Developing automatic mobility assistance systems for the safe navigation of visually impaired pedestrians in unstructured environments presents several complex challenges. The dynamic and unpredictable conditions and the diverse range of obstacles encountered by a pedestrian walking on an unstructured road add to the difficulties of automatic navigation. Another significant aspect of safe mobility assistance is real-time performance, which demands lightweight architectures and fast processing. In this paper, we propose an obstacle detection framework combining a road object detection model and a road anomaly detection model, using parallel processing for fast real-time performance. The models are based on convolutional neural network backbones, use transfer learning, and are trained on custom datasets manually collected in unstructured environments. The proposed system addresses the challenges of the complexities of the camera-based automatic navigation to detect obstacles and, based on their position and size, alerts the user via audio feedback.**

I. Introduction

In this paper, we propose a novel, deep learning and computer vision-based framework for automatically detecting outdoor objects and road anomalies for the safe mobility of visually impaired pedestrians in unstructured environments. We construct two novel image datasets of various kinds of road anomalies, such as puddles and potholes, and different road objects, such as trees, poles and signboards, that can cause distress and harm to visually impaired pedestrians, both captured manually from unstructured environments. We train object detection models on the curated datasets. Finally, we design a parallelly combined real-time system for fast and reliable detection of obstacles that notifies the user based on the risks posed by their surroundings.

Unstructured environments include road systems and cityscapes with inconsistent layouts. They are often characterised by an unpredictable road structure, absent pavements and sidewalks, and makeshift paths. These are often accompanied by poorly maintained roads and varying road conditions due to weather, ongoing construction work, irregular traffic, etc. Unstructured road systems are challenging to navigate because of the unpredictable and dynamic conditions, as shown in Figure 1. Such road systems are particularly inaccessible and risky for visually impaired pedestrians and can lead to accidents.

On the other hand, structured environments are well-planned, often comprising organised streets, designated zones for vehicles to pass, cycles, parking and pedestrian pathways, and an overall standardised infrastructure. Such environments are more accessible to navigate because of the clearly defined structure. Most existing mobility assistance systems are trained on data captured from structured environments because they are easier to work on and due to a lack of diverse and reliable datasets of unstructured environments. As a result, most of the mobility assistance systems fail in unstructured environments. This gives rise to the paramount need for datasets captured from unstructured environments and mobility assistance systems trained on such datasets to function in outdoor environments, irrespective of their structure.



Fig. 1: Unstructured roads with limited walkable area with (a) a pothole and two-wheeler, and (b) puddles and a tree.

Visually impaired pedestrians face numerous difficulties while navigating unstructured roads, posing significant challenges to their mobility and independence. Uneven surfaces like potholes, puddles or sudden depressions can cause tripping, stumbling, or falls, leading to injuries. Detecting them in advance is highly challenging without vision. Identifying parked or moving vehicles, poles, trees, signboards, etc. and navigating these obstructions can be difficult and pose a collision risk. Judging the distance and speed of moving vehicles without sight is challenging and can lead to severe accidents. The severity of these difficulties can vary depending on the individual's level of vision impairment, their mobility aids, and their familiarity with the environment. Weather conditions like rain and other circumstantial factors can further exacerbate these challenges. Overcoming these barriers requires improving the infrastructure design and assistive technologies that can adapt to unstructured environments.

[1] School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India aditya13_scs@jnu.ac.in, anwesh45_scs@jnu.ac.in, ayeshac@mail.jnu.ac.in
[*] Equal contibution
[2] Department of ECE, Delhi Technological University, Delhi, India s.indu@dce.ac.in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:33:05 UTC from IEEE Xplore. Restrictions apply.

Visually impaired people primarily rely on the traditional white cane for navigation. While undoubtedly instrumental, the white cane encounters significant limitations in unstructured environments, amplifying the existing navigational challenges. Firstly, the cane's limited reach restricts sensory input to the immediate ground, leaving potholes and broken roads undetected. Gauging the distance and size of more considerable obstacles like vehicles or poles proves challenging, hindering safe manoeuvring around them. Furthermore, the lack of defined paths or designated pedestrian zones on unstructured roads can throw navigation into disarray. Accurately judging the movement and speed of oncoming vehicles, which is crucial for safe road crossing, with a cane can be extremely risky. While the white cane remains a valuable tool, camera-based assistive systems are increasingly necessary to address its limitations on complex and unstructured terrains, ensuring more excellent safety and independence for visually impaired individuals.

While promising for visually impaired individuals, camera-based assistive systems come with their challenges. Obstacle detection is difficult amidst shadows, varying lighting, diverse backgrounds, and occlusion. Another major challenge is real-time processing and high computation demands, which is crucial for safety. In this work, we have considered road anomalies such as potholes, puddles, cracks, and patches, and road objects such as trees, poles, signboards, two-wheelers, vehicles and stairs for detection using two separate obstacle detection models, each of which comes with a different level of complexity due to their inherent structure, size and shape, as shown in Figure 1. We have addressed the problem of slow computation using a parallel processing framework and developed a real-time obstacle detection system that works efficiently in unstructured conditions.

II. Related Work

Blind and visually impaired people (BVIP) face significant challenges navigating urban environments, limiting their independence and mobility. [1] provides a comprehensive review of the overall state of research in outdoor navigation for BVIP. Some navigation systems for BVIP show promising results, but there are limitations in accuracy, reliability and user-friendliness. [2] reviews various assistive tools and technologies designed to help BVIP navigate their surroundings and categorises them into three main groups: Electronic Travel Aids that use sensors like ultrasound, LiDAR, or cameras, Position Locator Devices that use GPS and GIS technology for navigation, and Electronic Orientation Aids to understand the surrounding environment and spatial orientation using tools like talking maps, 3D tactile maps, and indoor positioning systems based on Bluetooth beacons.

Sensors and cameras have been used together to help visually impaired people navigate using sound or vibration to avoid collision[3]. However, more research is needed to see how well the system would work with larger groups of people and in different environments. [4] proposes an automated walking guide system designed explicitly as spectacles for visually impaired individuals. The system utilises three ultrasonic sensors to detect obstacles in the front, left, and right directions, along with a Convolutional Neural Network to identify potholes on the walking surface. The research is in its early stages, and further testing and development are needed before real-world implementation.

While sensors and wearable devices have made significant strides in assisting visually impaired individuals with mobility, they still face limitations due to their limited range of functioning, sensitivity to environmental challenges like rain, dust, etc., and difficulty differentiating between obstacles and harmless objects. Camera-based systems can adapt better to complex environments, the visual data allowing for a better understanding of surroundings, and are cost-effective and more accessible to carry, hence, user-friendly. [5] introduces a computer vision-based visual assistance system for the visually impaired, leveraging mobile edge computing and deep learning for advanced perception tasks like detecting and identifying various road elements, providing audio feedback, and utilising affordable hardware and mobile edge computing for broader accessibility.

[6] proposes an assistive device designed to aid visually impaired people with mobility and object recognition, combining camera and sensor data for richer information and robustness. However, processing camera data and performing object recognition require significant computational power, draining the battery life of devices and potentially limiting real-time performance. This can be a significant hurdle, especially for visually impaired users who rely on this technology for safe and independent navigation. [7] proposes a new framework to aid visually impaired people with outdoor mobility and provide rich auditory descriptions of their surroundings. It processes real-time video and multisensor data streams to identify the type of obstacles and the surrounding scene. It automatically generates the description of obstacles, their spatial information and corresponding instructions.

[8] leverages real-time image analysis to identify surroundings, transforming visual data into audio descriptions, in which users receive obstacle alerts, object recognition details, and spatial awareness through intuitive audio cues. The system personalises descriptions based on user preferences, granting control over the information received, but its accuracy relies on image clarity, and object misidentification can occur, posing safety risks. Additionally, processing power demands need to be addressed for wider adoption.

[9] proposes a framework by combining an object recognition module based on YOLOv-3 and a sensor-based obstacle detection system that works at chest level, waist level, knee level and foot level of the visually impaired individual. The information from both these modules is then fused to create a comprehensive understanding of the surroundings and translated into audio descriptions. [10] also uses the YOLOv-3 obstacle detection model to train on the Pascal VOC dataset and proposes an application that gives audio alerts for the detected obstacles. [11] emphasizes uncertainty awareness for robust obstacle detection and spatial understanding by

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:33:05 UTC from IEEE Xplore. Restrictions apply.

leveraging just a stereo camera system to predict salient regions in the image and combining it with 3D spatial information from the camera for robust obstacle assessment.

There has been remarkable progress in the development of assistive technologies for the visually impaired in recent times, with the evolution of deep learning and computer vision technologies, sensors and other perception technologies. In this paper, we aim to address the shortcomings of the existing works and propose a real-time obstacle detection architecture combining two deep learning-based models for the detection of road objects and road anomalies, especially in unstructured environments. We use parallel processing in the combined system to tackle the challenges of high computation time and lower detection accuracy. Moreover, the proposed system is trained on datasets manually captured from unstructured environments for reliable obstacle detection and safe navigation of visually impaired pedestrians on unstructured roads.

III. Proposed Work

In this section, we describe the pipeline of our proposed obstacle detection system in detail. The framework consists of two obstacle detection models trained on separate manually curated datasets of various classes of road anomalies and road objects. The collected video data is split into image frames, pre-processed and carefully annotated. Then, deep learning-based models with a CNN backbone are trained to perform object detection and classification as per the chosen classes. We combine these two models in different ways to create an efficient mobility assistance architecture that detects obstacles and warns the user about objects ahead or nearby road anomalies.

A. Custom Datasets

The performance and accuracy of deep learning models directly depend on the quality, diversity, and variability of the dataset on which they are trained. The unavailability of adequate, reliable datasets of various obstacles and road anomalies in unstructured environments captured from the perspective of visually impaired pedestrians gives rise to the need for tailoring a custom dataset. In this work, we have created a novel image dataset of obstacles and road anomalies commonly encountered on the road, which can cause distress and accidents for visually impaired pedestrians.

The custom dataset for road anomalies contains four classes: potholes, puddles, cracks, and patches, and the road obstacles dataset includes six classes of road obstacles: vehicles, two-wheelers, poles, signboards, trees, and stairs. The image frames in the datasets have been extracted from videos manually captured using mobile cameras while walking on pavements and lanes and from inside a vehicle while driving on various roads. The data is captured from multiple regions in urban settings, at different times of the day, and under various weather conditions to infuse a broad range of variability in terms of surrounding illumination, complex backgrounds and image quality. The original datasets contain over 3500 images, each with one or more object class instances.



Fig. 2: Samples of image frames from the custom datasets for road objects and road anomalies.

In real-life scenarios, the significant issues visually impaired people face in obstacle detection while walking arise due to difficulties of camera focusing and different angles of holding the camera. As a result, the captured image frames are often blurry and out of focus, worsened due to motion. These challenges can be significantly overcome by incorporating data augmentation steps, such as adding blur, flipping and rotating the frames up to an angle, differing the brightness and exposure, and so on. We have augmented the original dataset to obtain a more diverse dataset of over 15000 frames, which is robust to the mentioned challenges. Data augmentation also helps in addressing the problem of overfitting. It improves the model's generalisation upon training on previously unseen data, which is crucial, especially for unstructured environments due to unpredictable and changing conditions.

B. Data Preparation and Annotation

The captured videos are split into frames, and only the relevant frames containing helpful information for training are extracted for annotation. Null frames with backgrounds without any obstacle or anomaly of interest are also included, indicating a safe and clear path. Each frame may contain multiple instances of the same or different classes. The type of instances and their position in the frames should be chosen carefully to be annotated for optimal model performance in real-life scenarios. We have selected only the instances that can pose an obstacle or cause disturbance to a pedestrian, such as objects in the same lane as the pedestrian, mobile obstacles that are approaching or obstructing their path, or a pothole ahead, before proceeding to annotation.

Each frame is resized to a uniform dimension of 640x640 square pixels for even faster training. After manually marking them using bounding boxes and smart polygons, we carefully annotated each relevant instance in the pre-processed frames according to the ten corresponding classes. Correct and accurate annotation of instances is crucial for extracting features and training deep learning models on the dataset.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:33:05 UTC from IEEE Xplore. Restrictions apply.

C. Object Detection and Classification

In the proposed pipeline, we have performed object detection on images using a supervised deep learning model based on the YOLOv-8 (You Only Look Once) architecture to train on the curated datasets. YOLO [12] is a series of widespread object detection and image segmentation models known for its low cost of deployment, fast training and state-of-the-art object detection performance. YOLO uses an end-to-end Convolutional Neural Network consisting of 24 convolutional layers, followed by two fully connected layers. It uses a single-shot object detection algorithm that detects the location of objects in a single pass of the image, making it computationally efficient and suitable for real-time usage. YOLOv-8, the latest in the YOLO series, uses different augmentation methods while training, such as mosaic augmentation, and uses an anchor-free model for predicting the centre of an object directly. Additionally, YOLOv-8 has achieved extraordinary accuracy for object detection [13] according to the standard COCO (Common Objects in Context) [14] benchmark.

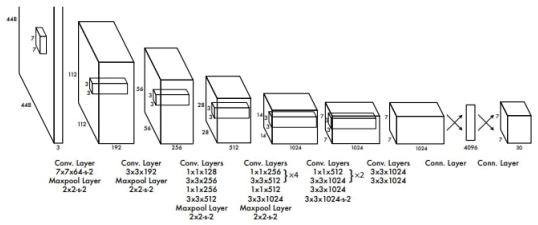


Fig. 3: YOLO Architecture [12]

The models are trained using transfer learning to achieve higher accuracy for all the different classes and then validated on validation datasets. The models are also tested separately on previously unseen video data. The two trained models can detect the obstacles and anomalies that appear in images and videos within bounding boxes and classify them as per the corresponding object class. The detection is accompanied by a confidence value. The object detection and classification on our novel datasets using YOLOv-8 show satisfactory results individually.

D. Proposed Parallel Architecture

Traditional object detection algorithms involve sequential information processing that becomes computationally expensive, especially with high-resolution images or videos. Parallel processing allows simultaneous data processing, significantly increasing speed and making real-time detection feasible. Our research employs two distinct models, each trained on a tailored dataset to tackle specific aspects of road safety. One model focuses on detecting road objects, while the other specialises in identifying road anomalies, both models addressing unique complexities inherent to their respective tasks. We evaluated three frameworks for object detection in input videos: (i) a sequential model architecture, (ii) a parallel model architecture, and (iii) a combined model architecture, comparing their performance in terms of both detection accuracy and execution time.

In the sequential architecture, a pre-processed frame is fed as input to one model, and the obtained output frame with detections is then provided as input to the other model. Then, in the parallel approach, we deploy our framework to test both models on the input frame in parallel using multi-threading to optimise the performance of both models and reduce execution time. The proposed parallel architecture is demonstrated in detail in Figure 4. In the combined approach, we combine both datasets into a single dataset of road objects and anomalies and train a single object detection model to detect the different classes at once.

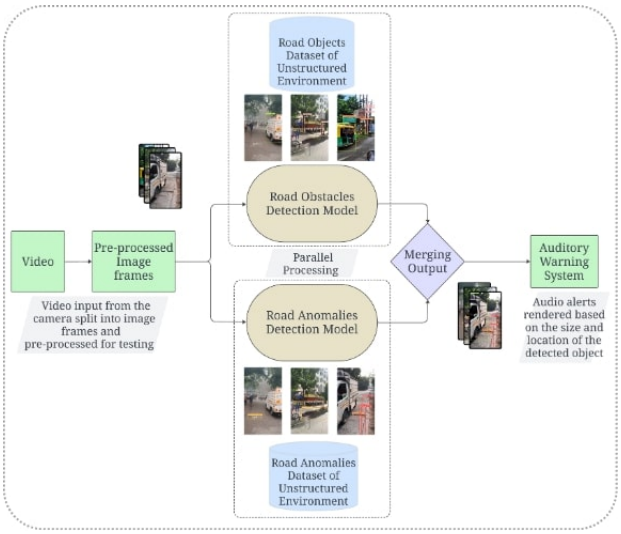


Fig. 4: The parallel architecture to test a video for road objects and road anomalies to achieve faster performance as compared to series architecture.

We observe and quantitatively elaborate in Section IV that among the three approaches, the parallel architecture gives optimal computation speed, along with accurate detections for all classes. The output detections of the models obtained in parallel are fused and integrated with the frame, resulting in faster and more efficient detection of obstacles. The user can switch between the parallel model framework and each of the distinct models at their convenience and depending on the environment they are navigating.

E. Audio Warning System

We integrate an audio-based alarm system with the parallel obstacle detection framework that notifies visually impaired pedestrians about the obstacle ahead based on its size and location, helping with safe navigation. If the area of the detected obstacle is more than 0.7 times the area of the frame, we empirically conclude that the obstacle is huge and close and can pose an immediate danger. Additionally, the centre of the detected obstacle being on the bottom two-thirds of the frame implies that the obstacle is close by and requires urgent attention. If the object's centre is near the centre of the image and the area is less than 0.7 times the area of the image, we

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:33:05 UTC from IEEE Xplore. Restrictions apply.

can deduce that the object is not obstructing the immediate path, which does not require immediate concern. Based on the threat level, the system sends either a high-priority critical or low-priority alert to trigger the user's attention. Figure 7 shows samples of high-priority and low-priority notifications issued based on the obstacles detected.

Visually impaired people rely heavily on ambient sounds to navigate their environment, gather information, and understand the surrounding activities. Moreover, a loud and repetitive alert can trigger a startled response and cause distress and disorientation in unfamiliar environments, leading to severe accidents. It is important to use calming and clear tones for notifications, and assistive devices should always have the option to regulate the volume and switch off the audio at the user's convenience.

IV. Experimentation Results and Discussion



Fig. 5: Low and inaccurate results of the combined model architecture with bounding box detections, class labels and confidence scores.

We have trained our models using object detection and transfer learning algorithms on our annotated datasets in an integrated GPU environment for faster training. The models have been configured in a parallel architecture on a system configured with AMD Ryzen 5 5600H and 8GB RAM, as shown in Figure 4. This architecture helped us achieve faster



Fig. 6: Accurate qualitative results of the proposed parallel architecture with bounding box detections, class labels and confidence scores.

TABLE I: A comparison of total run time and accuracy of different combinations of models.

Architecture	Duration (s)	Accuracy (F1-score)
Parallel Model	61.67	Object Accuracy - 0.88 @ 0.501 Anomalies Accuracy- 0.43 @ 0.286
Series Model	73.32	Object Accuracy - 0.88 @ 0.501 Anomalies Accuracy - 0.43 @ 0.286
Combined Model	49.49	0.54 @ 0.22

performance compared to a sequential implementation of the same models. We also compared the performance of the parallel architecture with a single object detection model trained on combining both datasets together. The parallel architecture gave a much higher accuracy compared to the single model, as shown in Table I.

Table I shows the overall time the different architectures take to run on the same video. The sequential model architecture that uses the models serially takes the longest time among the three. This is because both the models are inferenced one by one, which requires more computational time. The parallel model architecture takes slightly less time since it uses multiple threads simultaneously for computation.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:33:05 UTC from IEEE Xplore. Restrictions apply.

The combined model architecture in which a single model is trained on the combined dataset has a faster execution time but has the lowest accuracy due to the combination of complexities of ten different classes of objects. While the combined model fails, the parallel architecture takes slightly more computation time but achieves good detection and classification accuracy.

Figure 5 and 6 show the overall qualitative comparison of the same scenes using the parallel model architecture in Section III-D and the combined model architecture, respectively. From the images, we can infer that there is high complexity in detecting ten different classes of objects compared to using two different models to detect 4 & 6 classes each. The parallel model can detect all the various classes with high confidence because there are two different models doing two tasks individually. One detects all the road anomalies, and the second detects all the road obstacles, and the obtained results are merged, thus reducing the overall complexity.



(a) (b)

Fig. 7: Qualitative results with a textual warning based on the location of the road anomalies and obstacles. (a) Showing High Alert based on the location of the pothole and the size of the bounding box of the vehicle. (b) Showing Low Alert based on the location of the puddle and the vehicle.

The audio cues for the warning system have been changed here to textual cues, as shown in Figure 7 for better representation. The warning system works well in real time for all the anomalies and obstacles. It detects objects in each frame based on the overall image size and the bounding box location and size.

V. Conclusion

Unstructured environments present heightened challenges for the safe and independent mobility of visually impaired individuals. Our research underscores the immense potential of camera-based obstacle detection systems with audio alerts tailored to the complex landscapes and obstructions posed by unstructured environments. By leveraging obstacle detection algorithms, fast and real-time processing, and customised datasets, our findings demonstrate that the proposed obstacle detection framework can reliably identify and communicate obstacles and road anomalies, empowering individuals with spatial awareness and improved independent mobility.

Acknowledgement

This work has been supported by the CHANAKYA Fellowship from the IIIT Delhi iHub Anubhuti Foundation awarded to Aditya Kumar and the GATE PG Fellowship from AICTE awarded to Anwesha Chakravarty during their Master's studies in Computer Science and Technology from the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi. The authors sincerely thank both the organisations for their support.

References

[1] F. E. Z. El-Taher, A. Taha, J. Courtney and S. Mckeever, Sensors., A Systematic Review of Urban Navigation Systems for Visually Impaired People., 21(9), p.3103., 2021.

[2] K. Bineeth, IETE Technical Review., Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review., p.1, 2020.

[3] S. Kayukawa, T. Ishihara, H. Takagi, S. Morishima and C. Asakawa, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies., Guiding blind pedestrians in public spaces by understanding walking behavior of nearby pedestrians., 4(3), pp.1-22, 2020.

[4] M. M. Islam, M. S. Sadi, and T. Bräunl, IEEE Transactions on Medical Robotics and Bionics., Automated walking guide to enhance the mobility of visually impaired people., 2(3), pp.485-496, 2020.

[5] J. K. Mahendran, D. T. Barry, A. K. Nivedha and S. M. Bhandarkar, IEEE/CVF inproceedings of the Conference on Computer Vision and Pattern Recognition(CVPR)., Computer vision-based assistance system for the visually impaired using mobile edge artificial intelligence., (pp. 2418-2427), 2021.

[6] V. V. Meshram, K. Patil, V. A. Meshram and F. C. Shu, IEEE Transactions on Human-Machine Systems., An astute assistive device for mobility and object recognition for visually impaired people., 49(5), pp.449-460, 2019.

[7] W. Khan, A. Hussain, B. Khan, R. Nawaz and T. Baker, IEEE 12th International Conference on Developments in eSystems Engineering (DeSE)., Novel framework for outdoor mobility assistance and auditory display for visually impaired people., (pp. 984-989), October, 2019.

[8] A. G. Sareeka, K. Kirthika, M. R. Gowthame and V. Sucharitha, IEEE 2nd International Conference on Inventive Systems and Control (ICISC)., pseudoEye—Mobility assistance for visually impaired using image recognition., (pp. 174-178), January, 2018.

[9] S. Yadav, R. C. Joshi, M. K. Dutta, M. Kiac and P. Sikora, IEEE 43rd International Conference on Telecommunications and Signal Processing (TSP)., Fusion of object recognition and obstacle detection approach for assisting visually challenged person., (pp. 537-540), July, 2020.

[10] N. Rachburee and W. Punlumjeak, International Journal of Electrical & Computer Engineering., An assistive model of obstacle detection based on deep learning: YOLOv3 for visually impaired people., (2088-8708), 11(4), 2021.

[11] G. Dimas, D. E. Diamantis, P. Kalozoumis and D. K. Iakovidis, Sensors., Uncertainty-aware visual perception system for outdoor navigation of the visually challenged., 20(8), p.2385, 2020.

[12] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, IEEE Conference on Computer Vision and Pattern Recognition(CVPR)., You only look once: Unified, real-time object detection., (pp. 779-788), 2016.

[13] N. M. Krishna, R. Y. Reddy, M. S. C. Reddy, K. P. Madhav and G. Sudham, IEEE Third International Conference on Inventive Research in Computing Applications (ICIRCA)., Object detection and tracking using YOLO., (pp. 1-7). September, 2021.

[14] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár and C. L. Zitnick, Springer International inproceedings of Computer Vision–ECCV 2014: 13th European Conference Part V 13, Zurich, Switzerland., Microsoft coco: Common objects in context., (pp. 740-755), September 6-12, 2014.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:33:05 UTC from IEEE Xplore. Restrictions apply.

ISSN 0965-545X, Polymer Science, Series A, 2024. © Pleiades Publishing, Ltd., 2024.

NATURAL POLYMERS

Carboxymethyl Guar Gum-Based Bioactive and Biodegradable Film for Food Packaging

Tanvi Singh[a], Meenakshi Tanwar[a], and Rajinder K. Gupta[a,*]

[a] *Department of Chemistry, Delhi Technological University, Delhi, 110042 India*

**e-mail: rkg67ap@yahoo.com*

Received January 3, 2024; revised March 3, 2024; accepted March 10, 2024

Abstract—The study aimed to synthesize ascorbic acid (AA) loaded carboxymethyl guar gum (CMG) hydrogel based film cross linked by citric acid (CA). Bioactive films (CMG-CA and CMG-CA-AA) were synthesized by solvent casting method. The flexibility and mechanical strength of the bioactive film (CMG-CA-AA) were enhanced by the addition of glycerol and optimized by analyzing tensile strength. The structural, morphological and thermal analysis was done by FTIR, XRD, SEM and TGA. Further, the films were evaluated for its anti-oxidant, anti-microbial property and cytotoxicity. Barrier properties, biodegradability and fruit shelf life-enhancer capability were also assessed. The outcomes of the testings demonstrated favorable results and confirms the potential of CMG-CA-AA film in food packaging.

DOI: 10.1134/S0965545X24600388

INTRODUCTION

In recent years, plastic packaging has raised serious disposal and environmental challenges that endanger human health and development [1]. A significant amount of these plastic materials is used in food packaging [2]. Merely 9% of plastic waste is recycled globally, while 22% is mishandled [3]. It is vital to investigate innovative food packaging with varied functional features which can provide enhanced food product protection [4]. Consumers prefer the naturality of food and extended shelf life when it comes to packaged food. People would prefer one product over another if it has a longer shelf life [5]. Synthetic polymers or plastics have low biodegradability, minimal reuse and recycling options, and other negative environmental effects [6]. The utilization of bioactive packing materials as a substitute of plastic packaging may be the solution to such problems [7]. Natural polymer-based hydrogel films (bio-based films) as food packaging materials can serve as a great alternative [8]. Natural materials are emphasized and urged to be used instead of other options since they are environmentally beneficial [9]. Lipids, proteins, polysaccharides, or a mixture of such substances could be used to make bio-based films [10]. Carbohydrate polymers stand out among them as promising materials for food packaging due to their superior mechanical qualities (such as tensile strength, toughness, and elongation) [11].

In this context, carboxymethyl guar gum hydrogel based film (CMG-CA-AA) is being considered as a suitable substitute. The film was prepared from carboxymethyl guar gum cross-linked by citric acid, loaded with ascorbic acid, where glycerol is used as a plasticizer. All of the raw materials used in preparing CMG-CA-AA films are non-toxic and inexpensive, and the preparation process is simple and cost-effective.

Guar gum (GG) is a seed gum made from *Cyamopsis tetragonolobus* embryos of the Leguminosae family [12–14]. Carboxymethyl guar gum (CMG) is a derivative of GG, which is preferred over the pure form due to GG's low thermal stability, irregular hydration rate, decreased viscosity upon storage, and higher risk of microbial contamination [15]. Carboxymethylation is the chemical technique used here for attaching pendant carboxymethyl groups to pure GG [16].

Cross-linking is a typical method for enhancing the performance of polymers [17]. Citric acid, a polycarboxylic acid naturally found in citrus fruits like lemons, is a cost-effective and non-toxic example that has been utilized to enhance the characteristic properties of polymers [18, 19].

CMG cross-linked CA hydrogel films lack innate antioxidant and antimicrobial properties. AA was loaded in the bioactive film to incorporate these properties. Films possessing antimicrobial and antioxidant properties used in food packaging can improve the quality of food items and prevent microbial development [20, 21]. Ascorbic acid (AA) is a naturally occurring compound with strong polarity and significant antioxidant action found in many foods, including mangoes, oranges, lemons, blackberries, and vegetables [22, 23]. The mechanical strength of a material is a crucial factor to consider when creating packaging

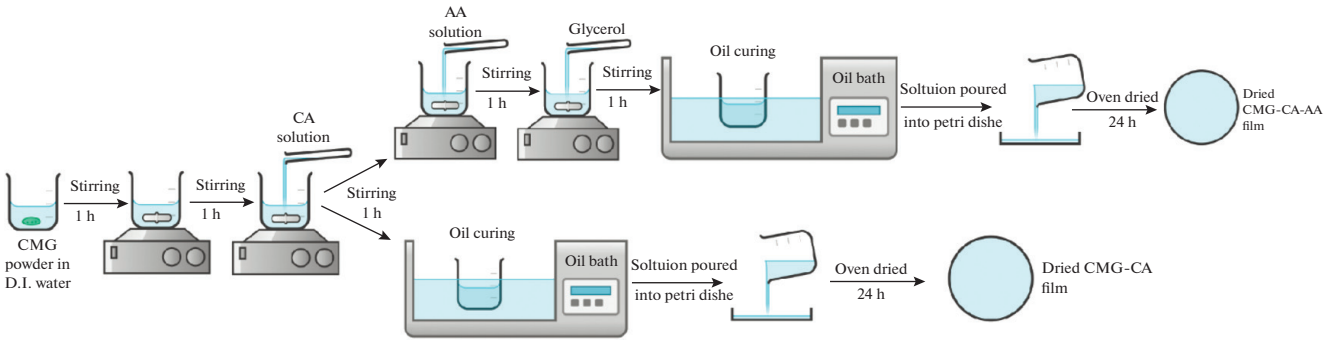


Fig. 1. Preparation method of (*1*) CMG-CA and (*2*) CMG-CA-AA films.

material. To improve this property of the prepared biofilm, a plasticizing agent, glycerol, was incorporated. Glycerol is a naturally occurring alcohol that is odorless, colorless, and slightly sweet to taste, and it was found to enhance the mechanical strength and elasticity of the films [24, 25].

In addition to being biodegradable, non-toxic, colorless, and odorless, the film also exhibits high barrier properties. Materials with high transmission barrier materials are highly suitable for providing package protection.

EXPERIMENTAL

Materials

CMG powder was purchased from the local market and CA was purchased from CDH, New Delhi. AA was purchased from CDH. Glycerol used in the preparation was from Fisher chemical, UK.

Preparation of CMG-CA-AA Films

Fabrication of CA cross-linked CMG bio active films was achieved through the solvent casting method (Fig. 1). Firstly, CMG 0.1% (w/v) was weighed and dissolved in distilled water to make a homogenized solution. 0.1% (w/v) CA dissolved in distilled water was added to the homogenized gum solution and stirred for 1 hour. 0.1% (w/v) AA dissolved in distilled water was added to the cross-linked CMG. After 1 h of stirring, 1 mL of glycerol was added and further stirred until a homogenized solution was obtained. Oil curing of the final solution was performed at a constant temperature of 140°C for 5–7 min. Controlled CMG-CA films were prepared by the same procedure, except the addition of AA. The prepared solution was transferred into Petri dishes and left to dry at 45°C for 24 h. After complete drying, the films were removed by peeling from the Petri dish.

Rheological Analysis

Rheology is an effective method for assessing the flow and deformation of matter by combining its principles with different properties [26]. Rheological studies of liquid CMG-CA-AA were performed on Anton Paar Modular Compact Rheometer 302 (MCR) at 25°C. Parallel plates of 40 mm were used for the testing. The constant shear rate for the steady shear rheology experiments varied between 0.1 and 500 s^{-1}, while the variable shear rate spans from 10 to 100 s^{-1}.

Mechanical Properties

The mechanical properties such as the thickness of films, maximum force, elongation at break (EB%), and tensile strength (TS) of the CMG-CA and CMG-CA-AA films were studied using, Universal Testing Machine (UTM), LLOYD LR 5K. The load range of the machine was 1–50 N with a 1000 mm extension range and 5 mm gauge length. The test and approach speed was 500 mm/min with no preload. A strip of 15 mm width and 150 mm length was cut from each film for mechanical testing. Each sample had three readings taken under air conditions, and the average result was taken into account in analyzing the mechanical characteristic. The *EB%* was calculated using the Eq. (1) [27]:

$$EB\% = \frac{E}{L} \times 100, \quad (1)$$

where E is the extent at which the film breaks and L is the initial length of the film.

The TS of the films were calculated using the Eq. (2) [28]:

$$TS = \frac{F}{A} \text{ (SI unit of TS is MPa)}, \quad (2)$$

where F is the force required to break the film and A is the cross-sectional area of the film.

Fourier Transform Infrared (FTIR) Spectroscopy

The FTIR spectra of CMG, CMG-CA and CMG-CA-AA were obtained through a Fourier transform infrared spectrophotometer (Perkin-Elmer model 2000). The range of spectra recorded was kept between 4000–400 cm^{-1} at room temperature [29].

X-ray Diffraction (XRD)

XRD analysis of CA, CMG, CMG-CA and CMG-CA-AA was performed using an Expert Pro MDR, Panalytical X-Ray diffractometer to get crucial information about its solid-state structure [30]. The scan rate was 1° per minute in the range of $2\theta = 5°$ to 80° using a copper diode with K_α radiation λ = 1.5406 Å.

Thermogravimetric Analysis (TGA)

Thermal degradation of CA, CMG, CMG-CA and CMG-CA-AA was determined by Thermo Gravimetric Analysis (TGA 4000, Perkin Elmer, USA). In a platinum crucible, approximately 4 mg of each sample was heated in a range of 25–800°C at a heating rate of 10 grad/min in N_2 gas atmosphere [31].

Scanning Electron Microscopy (SEM)

The morphology and physical properties of CMG-CA and CMG-CA-AA films were analyzed through a scanning electron microscope (SEM) (ZESIS EVO MA15). The samples were coated with gold to prevent electrostatic charge under a strong electron beam and low resolution during the investigation [32]. Film samples were observed at an accelerating voltage of 5.0 kV at 5000× and 10000× magnification.

Tests to Evaluate the Efficacy of the Hydrogel Film As a Potential Food Packaging Material

In vitro antibacterial activity. A disc diffusion susceptibility test was carried out to check the antimicrobial activity of the films [33]. The testing was done on agar plates for two different bacteria; gram-positive *Staphylococcus aureus* and gram-negative bacteria *Escherichia coli*. Overnight broth cultures of both bacteria were used for the assay. Under sterile conditions, the agar solution was poured and kept in Petri dishes to solidify. Bacterial inoculum was individually spread on the agar plates in an even manner. A disc of 1 cm diameter was cut from CMG-CA and CMG-CA-AA films and placed on agar plates with bacterial spread [34]. These plates were sealed and kept in an incubator at a temperature of 37°C for 24 h. The width of the inhibition zone (mm) around each sample was noted. Using a digital ruler, the sizes of the inhibition zones were calculated in millimeters (mm).

Antioxidant assay by DPPH radical scavenging. DPPH (2,2-diphenyl-1-picrylhydrazyl) is a free radical with an absorption maximum at 515 nm [35]. Antioxidants reduce the absorbance of DPPH. The samples' $DPPH^+$ free radical scavenging capacity was evaluated by using the $DPPH^+$ radical scavenging assay [36]. 1 mg $DPPH^+$ dissolved in 20 mL of methanol was prepared to be used as a radical solution. In methanol solution, DPPH gives a violet or purple color to the solution. The violet color fades to light yellow in the presence of an antioxidant. The antioxidant solution was prepared by mixing 1 mL of CMG-CA-AA homogeneous solution with 3 mL of methanol. Sample solutions were prepared with 1 mL $DPPH^+$ radical solution added to 4 mL antioxidant solution for the photometric assay [37]. At first, the absorbance of 1 mL $DPPH^+$ radical solution was recorded [using Ultraviolet–visible spectroscopy (Shimadzu UV spectrophotometer UV-1800)] as a blank, after that the $DPPH^+$ radical solution was added to the antioxidant solution. Then the readings were measured, one instantly and another at an interval of 30 min. The same procedure was repeated for all five different concentrations of CMG-CA-AA. The antioxidant activity was evaluated by analyzing and calculating the reduction in absorbance of $DPPH^+$ at different concentrations [38]. The following Eq. (3) was used to calculate the percentage $DPPH^+$ radical scavenging [39]:

$$DPPH^+ \text{ scavenging activity} = \left[1-\left(\frac{a-b}{c}\right)\right]\times 100, \quad (3)$$

where, a is the absorbance of sample solution, b is the absorbance of blank solution and c is the absorbance of $DPPH^+$.

A percentage higher than 85% indicates antioxidant activity [40].

Biodegradability test using burial method. The loaded bioactive film was subjected to degradation tests by the method of burring the sample in standard soil (ISO 17556 standard soil) for 25 days at room temperature and moisture controlled conditions. The sample was buried 2 cm under the soil. At intervals of 5 days, the buried film was taken out of the soil, gently washed with distilled water, dried and weight changes were noted [41]. The extent of degradation was evaluated by examining the changes in appearance, weight change and chemical changes through TGA. TGA was performed on a piece of bioactive film that had been buried in the soil for 15 days [42]. The results were compared with the TGA results of a freshly prepared bioactive film. The $\%D$ that is percentage weight loss

(at an interval of every 5 days) was calculated by Eq. (4) [43]:

$$\%D = \left(\frac{W - w}{W}\right) \times 100, \quad (4)$$

where W is the initial weight of the film and w is the weight after every 5 days.

Permeability Properties

Microbial penetration/microbial Barrier test. To determine the microbial penetration ability of the bioactive films, 2 vials of 5 mL were filled half (2.5 mL) with sterile nutrient agar broth. A circularly cut bioactive film (about 2 cm in radius) was fixed at the mouth of one of the vials [44]. The film was secured using rubber ties. The second vial was left uncovered and was used as a positive control for comparison. Both the vials were kept undisturbed for 24 h of incubation. After incubation, any cloudiness and turbidity observed in the nutrient agar broth in the vial was considered to be an indication of microbial contamination [45].

Water vapor permeability test. Using the desiccant method, the water vapor permeability of the bioactive films was evaluated. Two 5 mL vials were filled with 5 g of anhydrous $CaCl_2$ each. A circularly cut bioactive film (about 2 cm in radius) was fixed at the mouth of one of the vials (2.50 ± 0.05 m^2). The film was secured using parafilm and rubber ties. The second vial was left uncovered and was used as a reference for comparison [46]. The vials were put in a desiccator with a saturated NaCl solution at 37°C. The vials were weighed initially and after an interval of 24 h. The water vapor transmission rate (WVTR) can be calculated through Eq. (5):

$$WVTR = \frac{L}{A}\frac{\Delta m}{\Delta t}, \quad (5)$$

where Δm is the change in mass due to water vapor absorption (g), Δt is the time interval, L is the thickness (m) of the material through which water vapor passes and A is the exposed surface area of the material (m^2) [47].

Cytotoxicity evaluation by SRB assay. The cytotoxicity of the films on Hepatic (HepG2) cell line was analyzed by a Sulforhodamine B (SRB) assay. The cells were grown on 96-well plates for 24 h at 37°C with 5% CO_2 in Dulbecco's Modified Eagle's (DMEM) medium supplemented with 10% fetal bovine serum (FBS) and 1% antibiotic solution. Cells were treated with various doses of concentration the following day (0.075–1.25%). (In an insufficient medium, several concentrations were created) Tri-chloro acetic acid (TCA, 10%) was added to each well and incubated for an additional hour after being incubated for 24 h [48]. The plate was then rinsed with demineralized water (DM) water and allowed to air dry at room temperature. Each well received 0.04% SRB Solution, which was then added and allowed to sit for an hour. The plates were allowed to air-dry at room temperature after being incubated for 1 h. To remove any remaining unbound dye, they were then washed with 1% (v/v) acetic acid. In order to solubilize the protein-bound dye, tris base solution (pH 5) was added, agitated for 10 min on an orbital shaker, and then read at 510 nm in an Elisa plate reader (iMark, Biorad, USA).

Shelf life analysis (oranges). Packaging trials of shelf life on fresh fruit (orange) were done to examine the preservation capability of the bioactive films. Commercially, oranges are stored at a temperature of 3–9°C with 90–95% optimum humidity, to avoid spoilage and extend shelf life [49]. Inaccessibility to such optimum conditions results in spoilage and waste of food. The fabricated bioactive film is an attempt to provide sustainable packaging material with the least additional storage conditions. Three oranges were taken and placed in a separate petri dish. The first orange was left unpacked, i.e., without any film covering. Second, was packed with the CMG-CA film. Third was packed with the bioactive CMG-CA-AA film. Fruits were completely sealed in the films. All the samples were examined for 8 days. The samples were kept at room temperature in proper natural lighting [50]. Changes in the samples regarding spoilage and freshness were compared between the initial and last day (8th). These changes were analyzed to deduce a result regarding the preservation ability of the fabricated bioactive film (CMG-CA-AA) [51].

RESULTS AND DISCUSSION

Rheological Properties

Viscosity. Rheological analysis, clarifies the reaction between force, deformation, and time [52]. The experimental method of rheology is used to ascertain the rheological characteristics of materials [53]. Figure 2 shows the viscosity sweep of CMG-CA and CMG-CA-AA solutions. In comparison to the CMG-CA solution, the CMG-CA-AA solution was found to have a stronger shear thinning behavior [54]. The results suggest that with the introduction of AA into the hydrogel system a decrease in viscosity is observed. This could be as a result of the introduction of another acid, AA, which interrupted the intermolecular connections within the hydrogel system [55].

Mechanical Properties

The CMG-CA and CMG-CA-AA films were found to be 0.08 ± 0.005 and 0.03 ± 0.005 mm thick, respectively. Table 1 shows the results of mechanical testing for both films. Figure 3 shows the prepared hydrogel based films. The maximum load CMG-CA could hold was 10.60 N, while CMG-CA-AA could hold up to 39.90 N. The tensile strengths of CMG-CA and CMG-CA-AA were 7.20 and 29.08 MPa, respectively. This indicates that the bioactive film CMG-

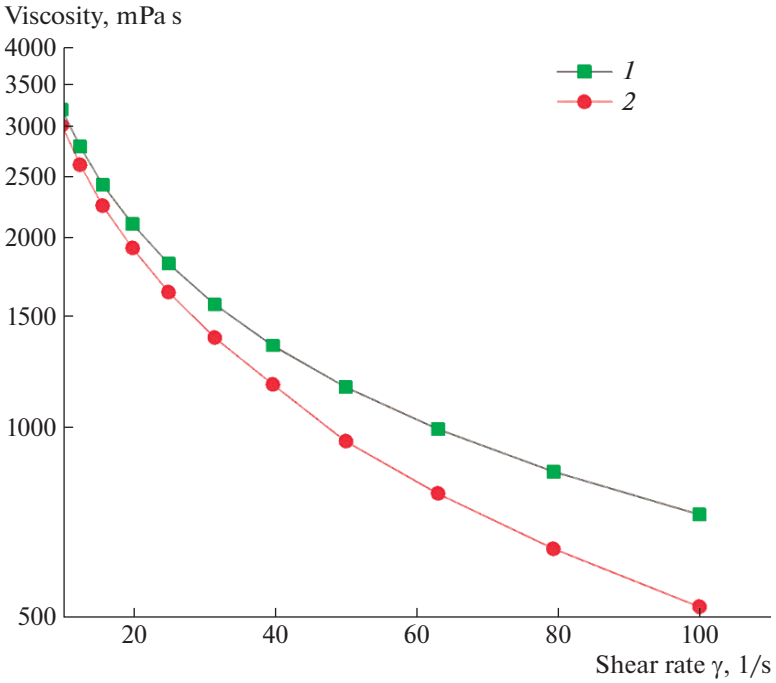


Fig. 2. Viscosity sweeps of (*1*) CMG-CA and (*2*) CMG-CA-AA solutions.

CA-AA has a higher tensile strength in comparison to CMG-CA, which proves that it can endure more force without breaking. The presence of glycerol and cross-linking of polymer chains in the bioactive films is the reason for their enhanced mechanical properties [56]. The results show the mechanical properties of the prepared bioactive film is greater than the controlled film. Such flexible and strong films could be used as a packaging film.

Fourier Transform Infrared (FTIR) Spectroscopy

FTIR spectroscopy was done to determine the functional groups of the films. Figure 4 shows the spectra of CMG, CMG-CA and CMG-CA-AA films. The vibrational stretching of –OH was shown by large, prominent peaks in the CMG spectrum between 3500 and 3000 cm^{-1}. The medium's peaks at 2852 and 2925 cm^{-1} demonstrated the asymmetric stretching of $-CH_2$ [57]. The –CO of ester group was said to be responsible for the highest peak presence at 1720 cm^{-1}. The C–O–C length of the glycoside link of CMG that resulted in the peak at 1023 cm^{-1}. The FTIR spectra of the cross-linked films CMG-CA and CMG-CA-AA showed the ester cross-links and carbonyl bond of free carboxylic acid at 1738 and 1726 cm^{-1} [58].

X-ray Diffraction (XRD) Analysis

The XRD data give important information on how composition affects crystal structure and shape. Figure 5 shows the X-ray diffractograms of CA, CMG, CMG-CA, and CMG-CA-AA. CMG powder show a broad peak at 19.08°. The CMG-CA and CMG-CA-AA films showed broad peaks at 2θ value 18.16° and 17.66°, respectively, indicating the amorphous structure of both films. Powder CMG shows a lower amorphous nature, indicated by the broad peak at 19.08°. Crosslinking via CA resulted in the substitution of the hydroxyl groups by the carboxymethyl groups in CMG-CA and CMG-CA-AA, caused an increase in amorphous structure. A reflex that is observed in the 40° region is also shown in the hydrogels of the gum (CMG-CA and CMG-CA-AA), confirming the

Table 1. Mechanical properties of the films

Film	Width, mm	Thickness, mm	Max force, N	Tensile strength, MPa	Elongation at break, %
CMG-CA	15.00	0.08 ± 0.005	10.60	7.20	47.50
CMG-CA-AA	15.00	0.03 ± 0.005	39.90	29.08	32.80

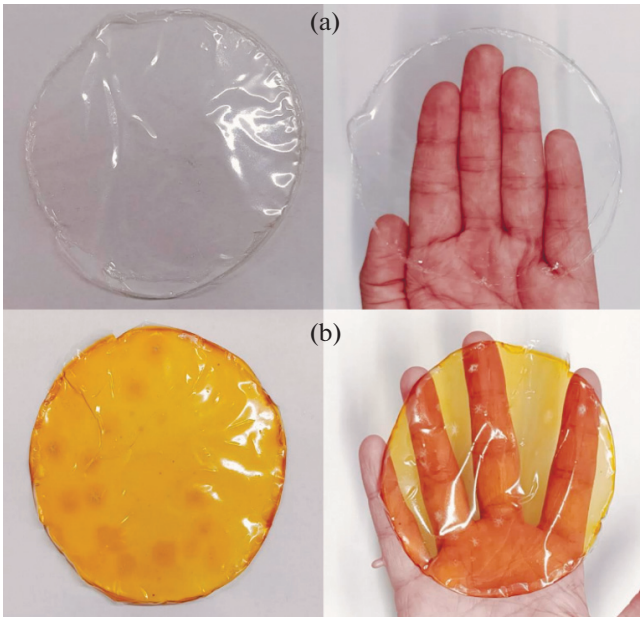


Fig. 3. (a) CMG-CA film, (b) CMG-CA-AA film.

cross-linking of the gum (CMG) and the citric acid. The breakage of hydrogen bonds of CMG crystals by cross linking increases the amorphous nature [59]. The moderately broader peak of CMG-CA-AA in comparison to CMG-CA at 18.36° suggests a smaller particle size which in turn suggests a higher amorphous nature of CMG-CA-AA [60].

Thermogravimetric Analysis (TGA)

The TGA curves shown in Fig. 6 of CA, CMG, CMG-CA, and CMG-CA-AA were studied to determine the thermal stability of the bioactive film. CMG-CA and CMG-CA-AA films showed a multi-step thermal degradation due to water evaporation, decomposition of CA and AA, breaking of polymer linkage bonds, and degradation of the polymer backbone. The 1st stage of mass loss is due to the evaporation of water molecules in all the samples tested [58].

In CA, the initial weight loss of 77.5% was observed between 172.89–237.55°C. CA completely degraded by the end of the analysis and no residual content was left.

In CMG gum 1st stage of degradation with 22.50 wt % loss was observed between 29.89–252.43°C. 2nd stage occurred between 252.43–555.88°C with a 50.42 wt % loss. And the final degradation was between 555.88–789.93°C with 23.53% and a residual content of 3.55 wt %.

In CMG-CA 1st stage of degradation was observed between 32.09– 229.99°C with a 11.29 wt % loss. 2nd stage was observed between 229.99–425.72°C(51.40%). The 3rd stage degradation at 425.72–780.62°C(32.47%). The residual content was of 4.84 wt %.

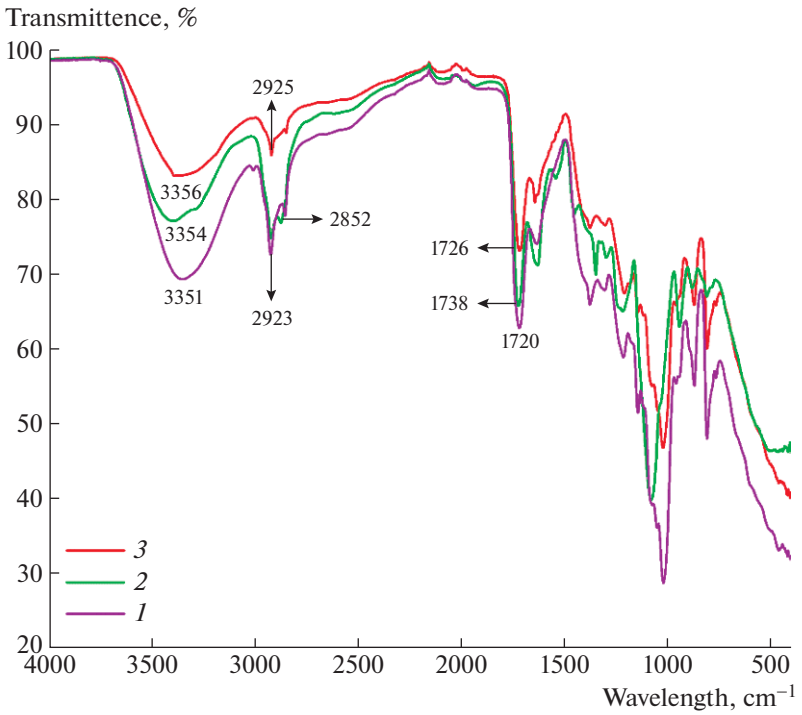


Fig. 4. FTIR graphs of (*1*) CMG, (*2*) CMG-CA and (*3*) CMG-CA-AA.

In CMG-CA-AA 1st stage of degradation was observed between 33.18–235.03°C with a 15.47 wt % loss. 2nd stage was observed between 235.03–502.86°C (53.16%). The 3rd stage degradation ranged between 502.86–615.33°C (11.19%). The 4th and final stage of decomposition was till 615.33–794.41°C (4.28%). The residual content was 15.90%.

In CMG-CA-AA, a slower thermal degradation occurred initially. Cross linking decreased the initial mass loss in both films in comparison to CMG. The initial temperature at which degradation starts was observed to be higher in cross linked films as compared to that of gum, indicating improved thermal stability. The residual content was found to be more in cross-linked films (CMG-CA and CMG-CA-AA) in comparison to CMG. CMG-CA-AA showed higher thermal stability than CMG-CA. The results of TGA give an idea about the thermal stability of the bioactive films which should be acknowledged while creating food packaging films.

Scanning Electron Microscopy (SEM)

The scanning electron micrographs of the surface morphology of CMG powder, CMG-CA and CMG-CA-AA films are presented in Fig. 7. CMG showed a globular structure. Both films exhibit small particle sizes and irregular particle shapes. SEM image analysis indicates that the CMG-CA film has compact surface with a few clusters of particles observed on the surface. On the other hand, CMG-CA-AA active films display a smoother, more compact and homogeneous surface with a patterned distribution of particle clusters which are favorable as a packaging film material [61]. Compared to the powder CMG, image of the CMG-CA-AA film at 1000× reveal smaller globular formations. This might have occurred due to cross-linking of the gum.

Tests to Evaluate the Efficacy of the Hydrogel Film As a Potential Food Packaging Material

In vitro antibacterial activity. Figure 8 illustrates the antibacterial activity of CMG-CA and CMG-CA-AA films. The images indicate that controlled CMG films did not exhibit any antibacterial activity, while CMG-CA-AA demonstrated good antibacterial activity. CMG-CA-AA showed a distinct zone of inhibition against both gram-positive (*Staphylococcus aureus*) and gram-negative (*Escherichia coli*) bacteria. The presence of AA contributed to the bacterial growth inhibition property exhibited by CMG-CA-AA. Incorporating 0.5 mg of AA into CMG-CA films produced a zone of inhibition of 15 mm (in diameter) against *Staphylococcus aureus* and 18 mm (in diameter) against *Escherichia coli*. Therefore, it can be inferred from the results that CMG-CA-AA films can prevent the growth of bacterial strains [62]. The antibacterial properties of CMG-CA-AA films can be significant for their application as a food packaging material [63].

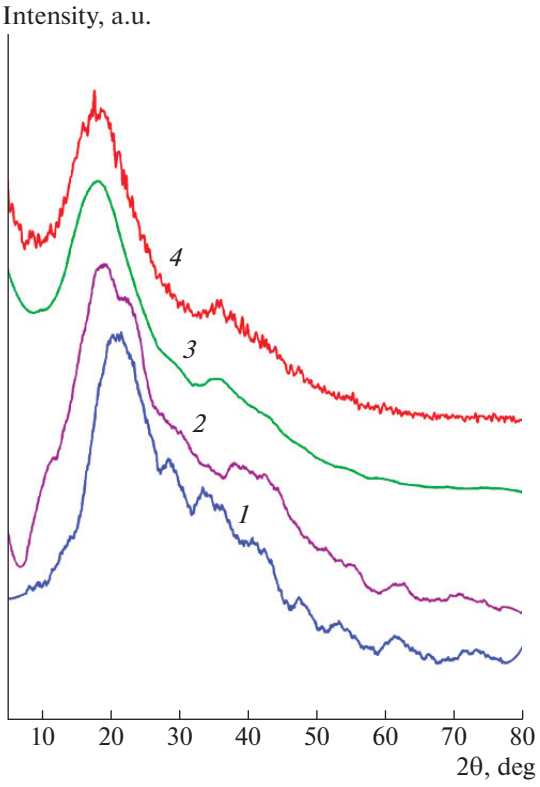


Fig. 5. XRD graphs of (*1*) CA, (*2*) CMG, (*3*) CMG-CA and (*4*) CMG-CA-AA.

Antioxidant assay by DPPH radical scavenging. Five different concentrations of CMG-CA-AA solution were tested to evaluate their antioxidant activity by analyzing the reduction of $DPPH^+$ radicals. The percentage radical scavenging activity of these sample solutions is shown in Table 2. The graphs of percentage $DPPH^+$ scavenging against concentration are presented in Fig. 9. The absorbance of $DPPH^+$ corresponding to the maximum wavelength ($\lambda_{max} = 515$ nm) was found to be 0.863. The scavenging percentage was highest (95%) at 500 μL/mL concentration and lowest (85.2%) at 100 μL/mL. Hence, higher scavenging was observed with an increase in concentration. The scavenging activity of the samples followed the order: 100 < 200 < 300 < 400 < 500 μL/mL. Antioxidants can reduce DPPH free radicals by donating hydrogen, which is essential to prevent the harmful role of free radicals in various disorders, including cancer [64]. The results suggest that all the loaded CMG-CA-AA samples exhibit antioxidant properties or radical scavenging activity due to their hydrogen donating or electron transfer ability. Antioxidant property in a food packaging film is a crucial requirement.

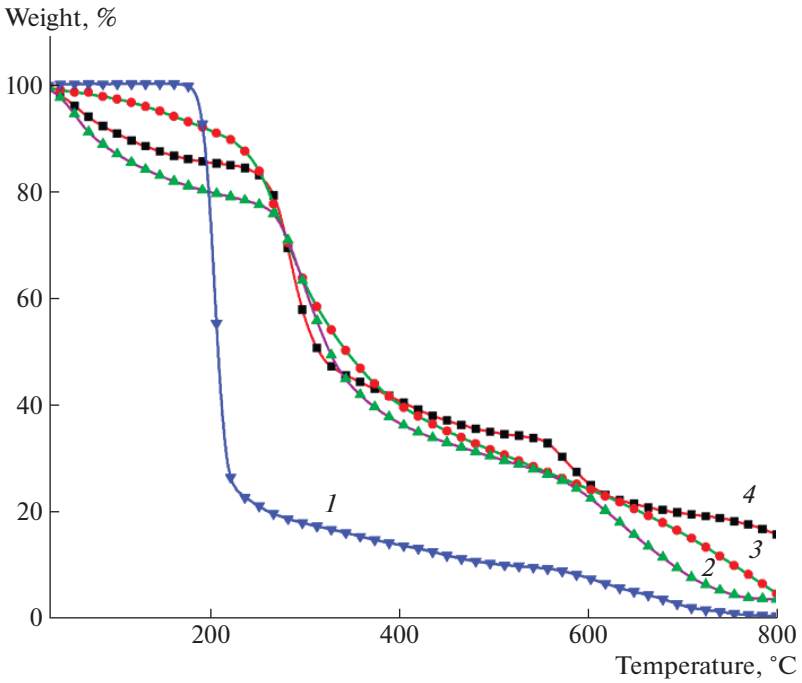


Fig. 6. TGA curves of (*1*) CA, (*2*) CMG, (*3*) CMG-CA and (*4*) CMG-CA-AA.

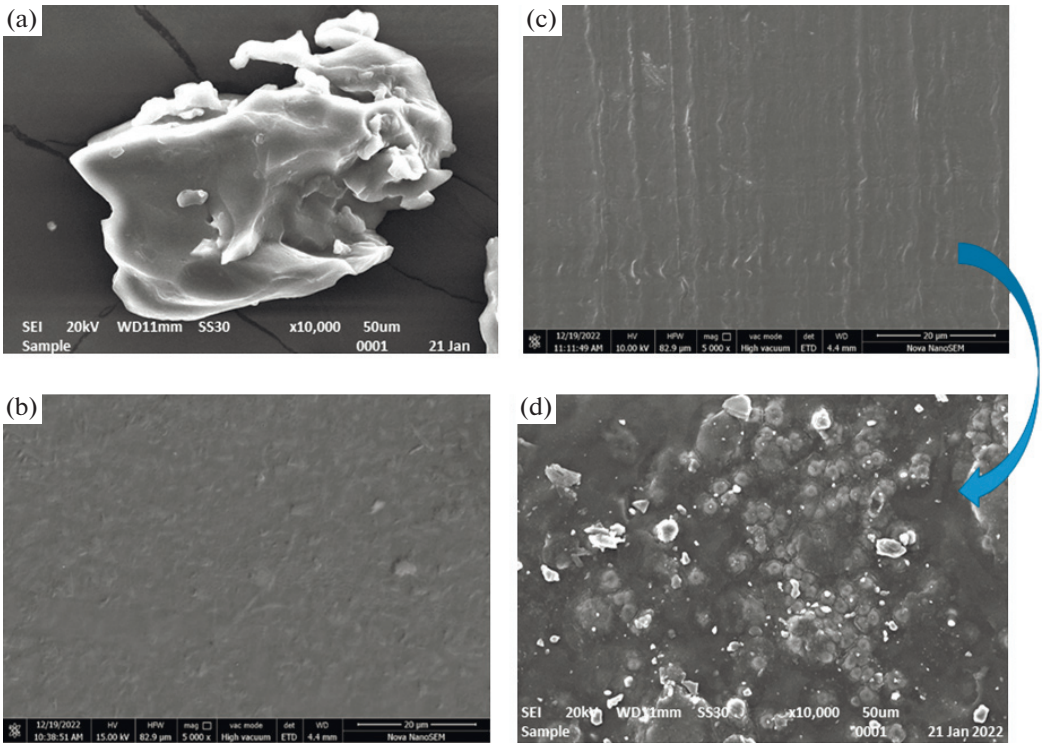


Fig. 7. Scanning electron microscope (SEM) images of (a) CMG powder at 10000×, (b) CMG- CA film at 5000×, (c) CMG-CA-AA film at 5000× and (d) CMG-CA-AA film at 10000×.

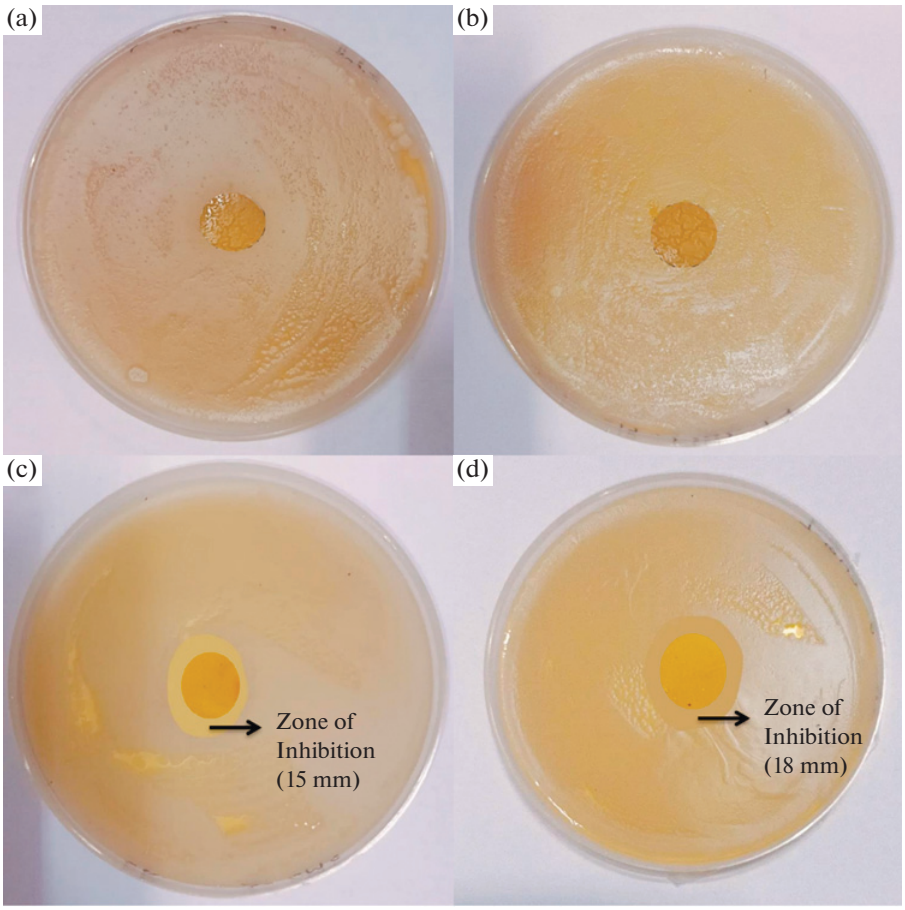


Fig. 8. In vitro antibacterial activity, (a) CMG-CA film in *Staphylococcus aureus*, (b) CMG-CA film in *Escherichia coli*, (c) CMG-CA-AA film in *Staphylococcus aureus*, (d) CMG-CA-AA film in *Escherichia coli*.

Biodegradability test using burial method. It was observed that the bioactive film started to degrade around the 5th day (week 1) after burial. The film started to develop cracks throughout and became softer. On reaching week 2 the film was seen to have broken down into a few pieces. In week 3 the film broke down into even smaller pieces. Most of the film had degraded by the starting of week 4 and only some of it could be seen in the soil with close observation. By the 25th day (week 4) no visible film particles could be observed and it was concluded that the majority of the film has degraded in the soil. At an interval of five days, the film was weighed to calculate the weight loss over the period of degradation [65]. The results are mentioned in Table 3.

During the evaluation, around 10th day (week 2) a fragment of the film was taken from the soil gently dusted, washed with distilled water and a thermal degradation analysis (TGA) was performed. The results of which was compared with a freshly prepared bioactive film (Fig. 10). CMG-CA-AA film showed a multi stage degradation while the degraded film has reduced to a single stage degradation. Shifts in the characteristic peaks and loss of residual contents in comparison to the initial film can be observed [66]. The results suggest that the components of the film have been broken down by the soil and is undergoing the process of natural decomposition. TGA results support the changes in the chemical structure of the degraded film. Figure 11 shows the degradation (change in

Table 2. Percentage radical scavenging activity

Concentration, μL/mL	Absorbance (after 30 min)	Percentage scavenging activity
100	0.128	85.2%
200	0.113	86.8%
300	0.103	88.1%
400	0.077	91.1%
500	0.044	95.0%

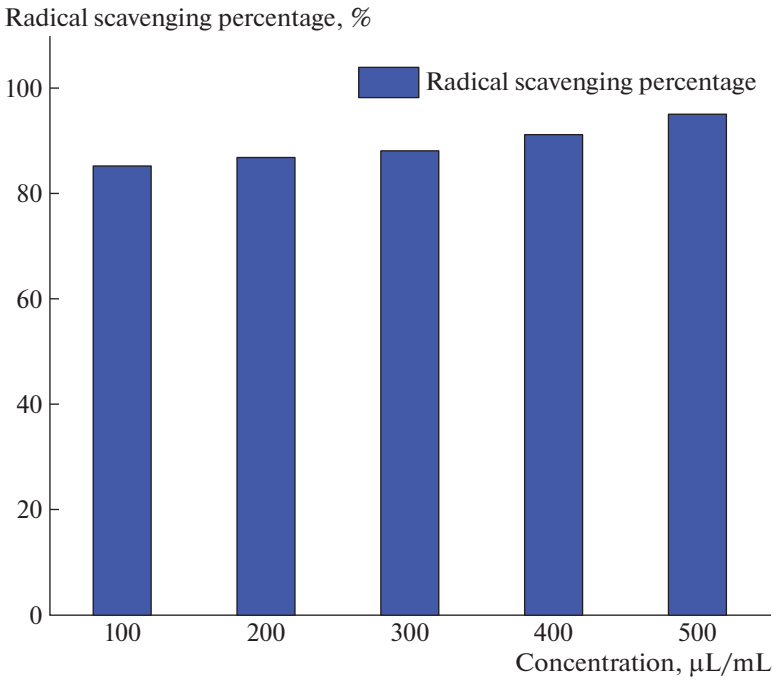


Fig. 9. DPPH+ scavenging percentage (antioxidant activity) graph of CMG-CA-AA.

shape and size) of the bioactive films over the course of 25 days. The results of the testing conclude that the prepared bioactive films are biodegradable and have significance in packaging film.

Permeability Properties

Microbial penetration/microbial Barrier test. After 24 hours, microbial growth in the uncovered vial was observed while the vial sealed with bioactive film showed no microbial growth. Figure 12 shows the observations of the test. White colored turbidity indicating microbial colonies can be observed in the nutrient medium of uncovered vial [67]. The nutrient medium of the bioactive film covered vial remains the same as in the beginning of the test that is a clear medium without any microbial growth. Through the findings of this test, it can be concluded that bioactive films inhibit the penetration of microbes through it. The presence of ascorbic acid in the bioactive films acts as an antimicrobial agent, which prevented the microbes from adhering to or growing on the surface of the film and also inhibiting the penetration of microbes through the film.

Table 3. Biodegradability assay: CMG-CA-AA film weight loss

Day	Initial weight, g	Final weight, g	Percentage weight loss, %
5	0.467	0.315	32.5%
10	0.315	0.204	35.2%
15	0.204	0.083	59.3%
20	0.083	0.019	77.1%
25	0.019	0.000	99%

Water vapor permeability test. There was a negligible change in the weight of the vial having bioactive film fixed on its mouth. Table 4 shows the change in weights of the vials used for testing. This indicates that the water vapor transmission rate (WVTR) of the bioactive film is negligible. For packaging films, high moist environment is considered to be unsuitable since moisture can promote degradation and contamination of food. The films will be effective in preventing a moist environment for the food [68]. The results suggest towards a favorable water permeable condition of the bioactive films as a food packaging material.

Cytotoxicity evaluation by SRB assay. Human hepatic cells (HepG2) were used as a model cell line to evaluate the cytocompatibility of CMG-CA-AA film. The cytotoxicity results for different concentrations are shown in Fig. 13. According to the results of the HepG2 cell culture experiment, the cells could adhere to the film surfaces, spread, and proliferate there [69]. Compared to the control, the cells on the CMG-CA-AA film showed a normal shape after incubation. It can be observed from Fig. 13 that all film samples of different concentrations exhibited cell viability values of more than 80%. This indicates the good biocompatibility of the films with HepG2 cells. Therefore, the

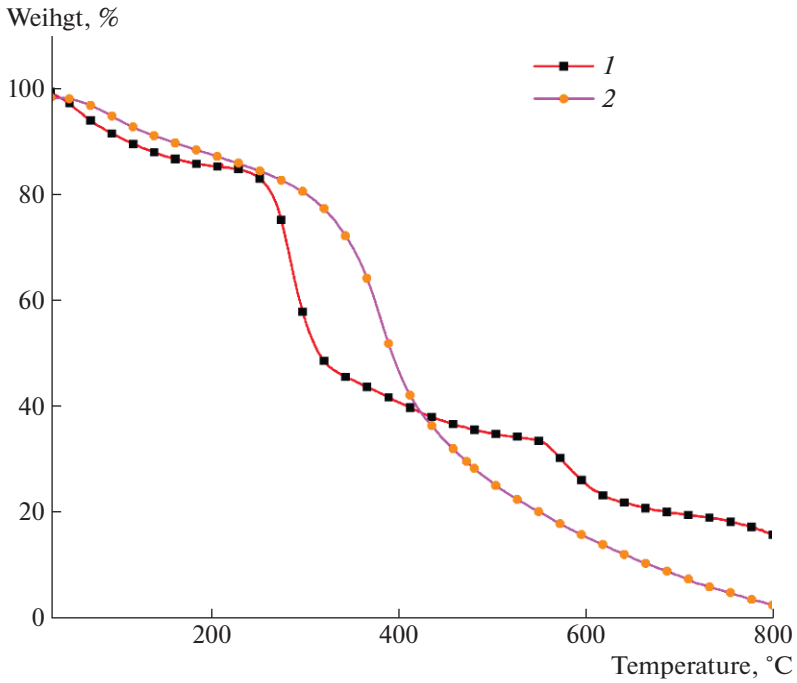


Fig. 10. TGA graph of (*1*) initial CMG-CA-AA and (*2*) the same film in degradation CMG-B.

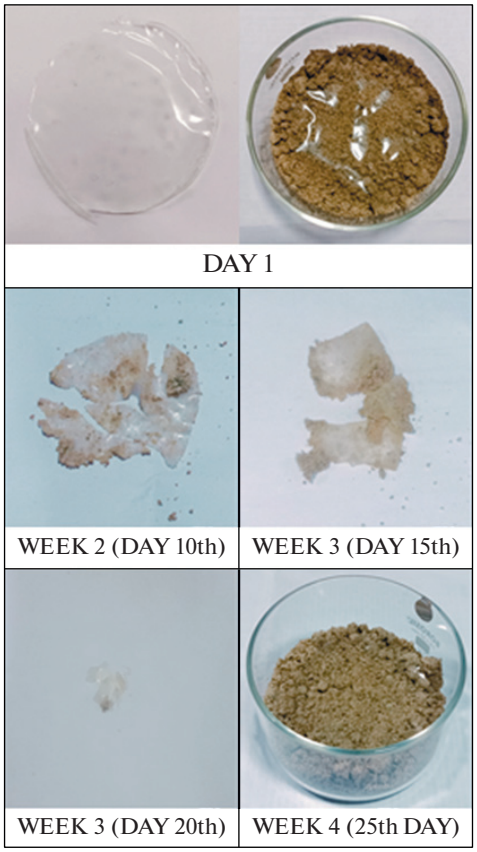


Fig. 11. Biodegradation of CMG-CA-AA film.

lower cytotoxicity of CMG-CA-AA film confirms the cytocompatibility of the bioactive film. The biocompatibility of the bioactive packaging film can be signified from this test.

Shelf life analysis (oranges). The food trials were conducted for 8 days to obtain accurate results [70]. Fruit trials are shown in Fig. 14. On day 1, all fruit samples appeared fresh, with an orange color and no microbial growth. On day 2, the oranges in the first and second Petri dishes began to ripen, while the third one remained unchanged. On day 3, the color of the oranges in the first two Petri dishes started to darken, but the third orange remained the same. On days 4 and 5, brown spots began to develop on the oranges in the first and second petri dishes, while there were no noticeable changes in the third orange. On days 6 and 7, the brown spots on the first and second oranges increased in size, and white microbial growth was observed on the first orange. The third orange started to ripen. On day 8, the first two oranges deteriorated significantly and were no longer suitable for consumption, while the third orange, covered with the bioactive film, looked similar to day 1 and was in perfect condition for consumption. These results indicate that the bioactive film successfully preserved the fruit and could be used as a packaging material.

CONCLUSIONS

The findings show that the CMG fabrication was effective in creating a bioactive film. In comparison to

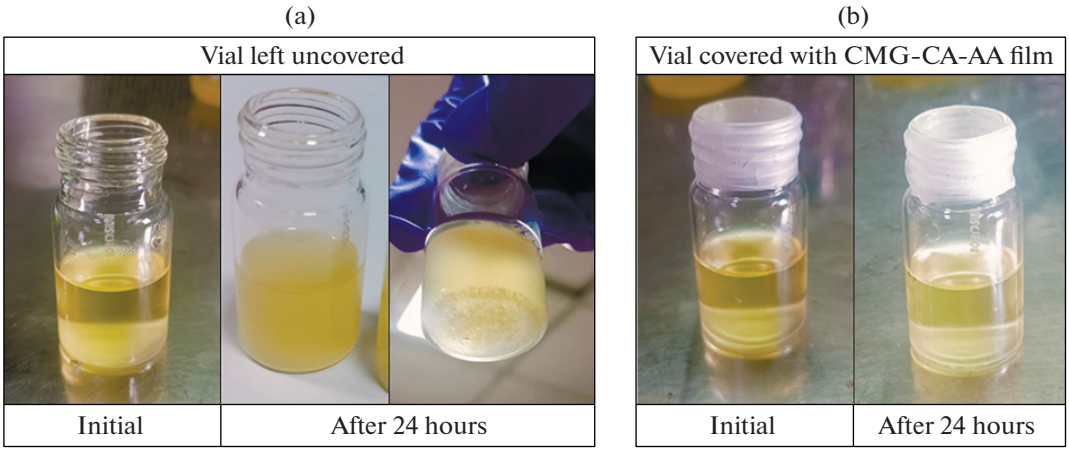


Fig. 12. Microbial barrier assay: (a) Vials left uncovered containing nutrient medium, (b) vials covered with CMG-CA-AA films containing nutrient medium.

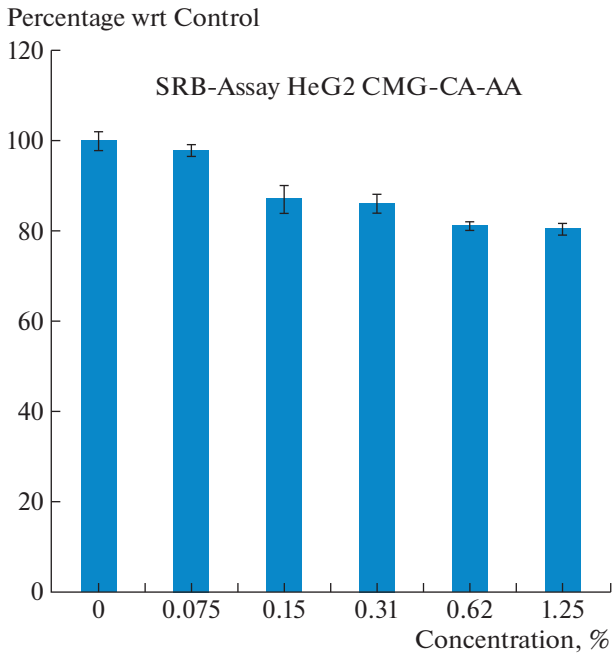


Fig. 13. Cytotoxicity evaluation, SRB assay-HepG2 of CMG-CA-AA at different concentrations.

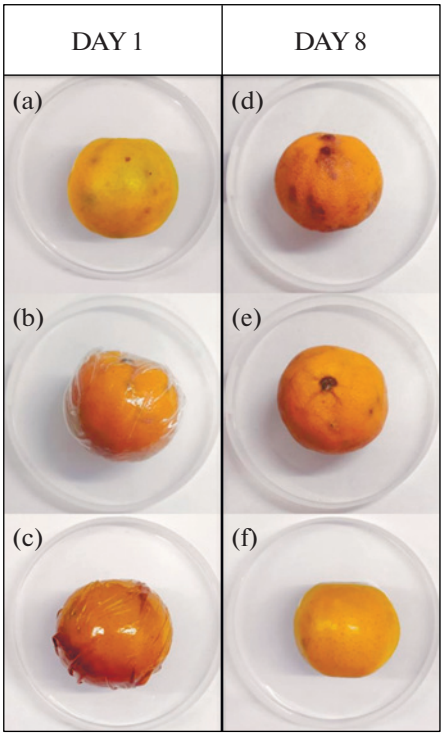


Fig. 14. Food sampling assay: (a) Uncovered orange at day 1, (b) Orange covered with CMG-CA film at day 1, (c) Orange covered with CMG-CA-AA film at day 1, (d) Appearance of uncovered orange after 8 days, (e) CMG-CA covered orange after 8 days and (f) CMG-CA-AA covered orange after 8 days.

the controlled film, the developed biofilm showed improved thermal stability and tensile strength. Granular particle was suggested by SEM pictures, and cross linking in the films was confirmed by FTIR measurements. The XRD results demonstrated the amorphous character of synthesized film. It also confirmed the cross-linking in CMC-CA-AA through a shift in the peaks. Effective suppression of both grams positive (*Staphylococcus aureus*) and gram negative (*Escherichia coli*) bacterial growth was demonstrated using CMG-CA-AA films. The films have antioxidant, biocompatible and biodegradable properties. The bioactive films appeared to be an effective barrier against microbial growth and water vapor based on studies of their barrier property analysis. The findings on food packaging demonstrated its usefulness as a packaging material by prolonging the freshness of fruits and extending their shelf life. Overall, CMG-CA-AA bio-

Table 4. Water vapor permeability test—change in weight

Sample	Weight of vial+$CaCl_2$, g (initial)	Weight of vial+$CaCl_2$, g (after 24 h)
Bioactive film fixed vial	15.384	15.385
Reference vial	14.822	15.097

active films have the potential to explore its application in food packaging.

ACKNOWLEDGMENTS

We are thankful to Aakaar Biotechnologies Private Limited for carrying out SRB assay of the samples. We are thankful to UFlex Private Limited for mechanical strength testing.

FUNDING

We are grateful to DTU for the financial and research support.

CONFLICT OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. E. J. North and R. U. Halden, Rev. Environ Health **28**, 1 (2013).
2. C. Y. Barlow and D. C. Morgan, Resour., Conserv. Recycl. **78**, 74 (2013).
3. L. S. D.-Hoffman, J. L. Lane, T. Grant, S. Pratt, P. A. Lant, and B. Laycock, J. Clean. Prod. **180**, 325 (2018).
4. S. B. Borrelle, B. M. Hamilton, and B. L. Clark, Science **369**, 1515 (2020).
5. J. P. Smith, D. P. Daifas, W. E.-Khoury, J. Koukoutsis, and A. E.-Khoury, Crit. Rev. Food Sci. Nutr. **44**, 19 (2004).
6. Z. Dorota, B. Bilska, K. M.-Łukasiak, A. Łepecka, M. Trząskowska, K. N.-Skocińska, M. Tomaszewska, A. Szydłowska, and D. K.-Krajewska, Int. J. Environ. Res. Public Health **17**, 1632 (2020).
7. E. Kabir, R. Kaur, J. Lee, K. H. Kim, and E. E. Kwon, J. Clean. Prod. **258**, 120536 (2020).
8. M. Asgher, S. A. Qamar, M. Bilal, and H. M. N. Iqbal, Food Res. Int. **137**, 109625 (2020).
9. J. Wang, M. Euring, K. Ostendorf, K. Zhang, J. Bioresour. Bioprod. **7**, 1 (2022).
10. Z. Yuan, B. Li, C. Li, Y. Xu, Y. Luo, D. Liang, and C. Huang, Foods **10**, 1845 (2021).
11. U. Amin, M. U. Khan, Y. Majeed, M. Rebezov, M. Khayrullin, E. Bobkova, M. A. Shariati, I. M. Chung, M. Thiruvengadam, Int. J. Biol. Macromol. **183**, 2184 (2021).
12. H. Prabhanjan, M. M. Gharia, and H. C. Srivastava, Carbohydr. Polym. **11**, 279 (1989).
13. G. Sharma, S. Sharma, A. Kumar, A. H. Al-Muhtaseb, M. Naushad, A. A. Ghfar, G. T. Mola, and F. J. Stadler, Carbohydr. Polym. **199**, 534 (2018).
14. D. Prem, S. Singh, P. P. Gupta, J. Singh, and S. P. S. Kadyan, Plant. Cell Tissue Organ. Cult. **80**, 209 (2005).
15. Y. Cheng, K. M. Brown, and R. K. Prud'homme, Biomacromolecules **3**, 456 (2002).
16. V. P. Chakka and T. Zhou, Int. J. Biol. Macromol. **165**, 2425 (2020).
17. R. Salihu, S. I. A. Razak, N. A. Zawawi, M. R. A. Kadir, N. I. Ismail, N. Jusoh, M. R. Mohamad, and N. H. M. Nayan, Eur. Polym. J. **146**, 110271 (2021).
18. C. Q. Yang, X. Wang, and I.-S. Kang, Text. Res. J. **67**, 334(1997).
19. C. Q. Yang and B. A. K. Andrews, J. Appl. Polym. Sci. **43**, 1609 (1991).
20. R. Sharma, S. M. Jafari, and S. Sharma, Food Control **112**, 107086 (2020).
21. Y. C. Yang, X. W. Mei, Y. J. Hu, L. Y. Su, J. Bian, M. F. Li, F. Peng, and R. C. Sun, Int. J. Biol. Macromol. **134**, 122–130 (2019).
22. T. Huang, Y. Qian, J. Wei, and C. Zhou, Polymers **11**, 560 (2019).
23. H. Baniasadi, Z. Madani, R. Ajdary, O. J. Rojas, and J. Seppälä, Mater. Sci. Eng., C **130**, 112424 (2021).
24. U. Shah, F. Naqash, A. Gani, and F. A. Masoodi, Compr. Rev. Food Sci. Food Saf. **15**, 568 (2016).
25. L. Li, S. Guan, L. Yang, X. Qin, and W. Feng, Chem. Res. Chin. Univ. **34**, 311 (2018).
26. G. Sun, T. Liang, W. Tan, and L. Wang, Food Hydrocoll. **85**, 61 (2018).
27. A. K. Patel, P. Michaud, H. Baynast, M. Grédiac, and J. D. Mathia, J. Appl. Polym. Sci. **127**, 3869 (2012).
28. B. Singh, S. Sharma, and A. Dhiman, Int. J. Pharm. **457**, 82 (2013).
29. V. E. Bosio, V. Machain, A. G. López, I. O. P. D. Berti, S. G. Marchetti, M. Mechetti and G. R. Castro, Appl. Biochem. Biotechnol. **167**, 1365 (2012).
30. M. Kalachaveedu, P. Jenifer, R. Pandian, and G. Arumugam, Int. J. Biol. Macromol. **148**, 737 (2020).
31. S. C. Barros, A. A. Silva, D. B. Costa, C. M. Costa, S. L. Méndez, M. N. T. Maciavello, J. L. G. Ribelles, F. Sentanin, A. Pawlicka and M. M. Silva, Cellulose **22**, 1911 (2015).
32. Q. S. Zhao, Q. X. Ji, K. Xing, X. Y. Li, C. S. Liu, and X. G. Chen, Carbohydr. Polym. **76**, 410 (2009).
33. E. A. Bursali, S. Coskun, M. Kizil, and M. Yurdakoc, Carbohydr. Polym. **83**, 1377 (2011).

34. V. Thomas, M. M. Yallapu, B. Sreedhar, and S. K. Bajpai, J. Colloid. Interface Sci. **315**, 389 (2007).
35. V. B. K. Schlesier, M. Harwat, and R. Bitsch, Free Radical Res. **36**, 177 (2009).
36. O. P. Sharma and T. K. Bhat, Food Chem. **113**, 1202 (2009).
37. A. L. Dawidowicz, D. Wianowska, and M. Olszowy, Food Chem. **131**, 1037 (2012).
38. W. B. Williams, M. E. Cuvelier, and C. Berset, LWT–Food Sci. Technol. **28**, 25 (1995).
39. C. Desmarchelier, M. J. N. Bermudez, J. Coussio, G. Ciccia, and A. Boveris, Int. J. Pharmacogn. **35**, (2), 116 (2008).
40. M. S. Blois, Nature **181**, 1199 (1958).
41. K. Sharma, V. Kumar, B. S. Kaith, V. Kumar, S. Som, S. Kalia, and H. C. Swart, RSC Adv. **4**, 25637 (2014).
42. H. L. Abd El-Mohdy and S. Ghanem, J. Polym. Res. **16**, 1 (2008).
43. W. Phetwarotai, P. Potiyaraj, and D. Aht-Ong, J. Polym. Environ. **21**, 95 (2013).
44. N. Roy, N. Saha, and P. Saha, Soft Mater. **8**, 338 (2010).
45. F. Yañez, J. L. Gomez-Amoza, B. Magariños, A. Concheiro, and C. Alvarez-Lorenzo, J. Membr. Sci. **365**, 248 (2010).
46. J. S. Gonzalez, L. N. Ludueña, A. Ponce, and V. A. Alvarez, Mater. Sci. Eng., C **34**, 5 (2014).
47. R. Singh and D. Singh, J. Mater. Sci. Mater. Med. **23**, 2649 (2012).
48. H. Qanash, R. Yahya, M. M. Bakri, Ab. S. Bazaid, S. Qanash, Ab. F. Shater and T. M. Abdelghany, Sci. Rep. **12**, 5914 (2022).
49. C. Chen, N. Cai, J. Chen, X. Peng, and C. Wan, Coatings **8**, 445 (2018).
50. M. Kaur and D. Santhiya, J. Appl. Polym. Sci. **138**, 49832 (2021).
51. S. D'Aquino, A. Piga, M. Agabbio, and T. G. McCollum, Postharvest. Biol. Technol. **14**, 107 (1998).
52. C. Wu, J. Tian, S. Li, T. Wu, Y. Hu, S. Chen, T. Sugawara, and X. Ye, Carbohydr. Polym. **146**, 10 (2016).
53. P. F. G. Banfill, Constr. Build. Mater. **8**, 43 (1994).
54. R. Yadav and R. Purwar, Polym. Test. **93**, 106916 (2021).
55. S. Jabeen, M. Maswal, O. A. Chat, G. M. Rather, and A. A. Dar, Colloid. Surf., B **139**, 211 (2016).
56. B. Singh, J. Singh, and Rajneesh, Carbohydr. Polym. **2**, 100058 (2021).
57. K. K. Mali, S. C. Dhawale, and R. J. Dias, Int. J. Biol. Macromol. **105**, 463(2017).
58. D. Mudgil, S. Barak, and B. S. Khatkar, Int. J. Biol. Macromol. **50**, 1035 (2012).
59. H. Daud, A. Ghani, D. N. Iqbal, N. Ahmad, S. Nazir, M. J. Muhammad, E. A. Hussain, A. Nazir, and M. Iqbal, Arab. J. Chem. **14**, 103111 (2021).
60. H. Chen, H. Wang, B. Li, B. Feng, X. He, W. Fu, H. Yuan, and Z. Xu, RSC Adv. **8**, 16910 (2018).
61. V. Sharma, P. Patnaik, K. Senthilguru, S. K. Nayak, I. Syed, V. K. Singh, P. Sarkar, G. Thakur, and K. Pal, Chem. Pap. **72**, 2101 (2018).
62. J. Bakht, H Ali, M. A. Khan, A. Khan, M. Saeed, M. Shafi, A. Islam, and M. Tayyab, Afr. J. Biotechnol. **10**, 19825 (2011).
63. R. Verghese, S. Mathew, and A. David, Int. J. Curr. Res. Med. Sci. **3**, 88 (2017).
64. E. M. Njoya, in *Cancer: Oxidative Stress and Dietary Antioxidant*, 2nd ed., Ed. by V. R. Preedy and V. B. Patel (Academic Press, 2021). pp. 349–357
65. G. M. E.-Villegas, G. Morselli, M. J. A. Oliveira, G. G.-Pérez, and A. B. Lugão, Polym. Bull. **77**, 4147 (2020).
66. R. Mohan, C. Sanpitakseree, A. V. Desai, S. E. Sevgen, C. M. Schroeder, and P. J. A. Kenis, RSC Adv. **5**, 35211 (2015).
67. N. Peng, Y. Wang, Q. Ye, L. Liang, Y. An, Q. Li, and C. Chang, Carbohydr. Polym. **137**, 59 (2016).
68. S. Guilbert, N. Gontard, and B. Cuq, Packag. Technol. Sci. **8**, 339 (1995).
69. V. Saravia and J. L. T.-Herrera, Microsc. Res. Tech. **72**, 957 (2009).
70. T. S. Parreidt, K. Müller, and M. Schmid, Foods **7**, 170 (2018).

Publisher's Note. Pleiades Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2024 5th International Conference for Emerging Technology (INCET)
Karnataka, India. May 24-26, 2024

Classification of Banana Plant Leaves Based on Nutrient Deficiency Using Vision Transformer

Sumit Kumar Sharma
Department of Information Technology
Delhi Technological University
Delhi, India
sumitkumarsharma_23phdit03@dtu.ac.in

Prof. Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
Delhi, India
dinesh@dtu.ac.in

2024 5th International Conference for Emerging Technology (INCET) | 979-8-3503-6115-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/INCET61516.2024.10593120

***Abstract*— Banana is one of major fruit crop consumed in India and its production is merely impacted over the year by many factors where malnutrition is one of the key causes which impact banana plants growth and production of banana. Identification of particular type of malnutrition in early stage avoids the diseases and lower the impact of that particular malnutrition on the plant. Traditional monitoring of plants requires intensive time and relies on humans to judge the type of malnutrition suffered a plant, whereas computer vision techniques currently using in similar tasks such as diseases detection, computer vision also performs well on the discussed task which discloses in this study here we proposes a vision transformers-based solution to classify the leaves based on the particular type of malnutrition it has. Vision transformer models have low inductive biases which makes them not to focus on local regions but also consider global patterns in image which makes to produce better results over images taken from real world crop fields. Due to which classification task able to classify with accuracy of 90 percentage on the benchmark dataset which is taken from Maland university of India in Karnataka, which includes 7000 plus images of banana plant leaves when leaves are under deficiency of 8 unique nutrition.**

Keywords*— nutrient deficiency, computer vision, Image classification, vision transformers

I. INTRODUCTION

Banana is one of the important majorly grown fruit in India, plants having different stages which includes consistently monitoring and care, lack of which introduces diseases or reduction in production [1]. And managing using existing methods requires huge time and cost as it is completely depend upon human labor which is not effective in terms of time and cost, whereas artificial intelligence able to solve similar kind of problems such as diseases detection where computer vision detects particular type of diseases on plant leave and helps to protect plant even in early stages of particular diseases associated with plant leaves [2]. In similar to diseases detection image processing is used to classify the leaves based on the deficiency of nutrient [4].There are various different types of image classification models which able to perform classification tasks over input images and classified them based on particular features like here classification of banana leaves performed based on the type of nutrient which a plant having deficiency, for classification tasks several benchmarking models were exist such as LeNet which was released in 1998 which uses pooling layers with smaller number of convolution layers[10]. In 2012 AlexNet another image classification model with introduces usage of ReLU activation, drop and level up the accuracy of image classifications tasks [11]. After that in 2014 VGGNet with uses small 3X3 filters for feature extraction with deep architecture based solution which surpass the AlexNet in image classification task over benchmarking datasets [12]. In same year GooLeNet popularly known as inception lunched which surpasses even VGGNet in some areas of image classification task as it enables extraction of features at multiple scales with reducing in the computational efforts [13]. Then in 2015 ResNEt introduces for image classification task which solves the problem of vanishing gradient, it was basically residual learning architecture with deep networks [14].Then in next year which is 2016 Densenet comes out which uses connected layers where each layers receives input from all the existing previous layers which promotes feature reuse, and it surpass the ResNEt in classification over benchmarking dataset [15]. After that in year 2017 NasNet and mobileNEt were introduced which perform classification task more accurate and even number of dense layers were also increases as we were moving one to another models [16, 17]. After 2 year another model which is EfficentNEt in 2019 was introduces which basically starting of balancing the model size as till now model gets deeper and deeper to generate better results [18]. These all are CNN based image classification models but vision transformers surpass the image classification task over the discussed CNN models as it captures the relationships between long range and different patches of input image which makes classification better over the benchmarking dataset [19]. Popular classifications models discussed above majority of them use ImageNet dataset [20] as benchmarking dataset for classification task. Motivation of the study is that nutrient deficiency detection in plants is requirement of current era where AgriTech sector entering into unmanned farming, The model must be in lesser resource intensive which makes it consumable in form of simpler API used for real time apps with lower response time and latency rates, transformers looks best solution potentially, as vision transformers attains excellent results compared to generally using convolutional networks while requiring substantially fewer computational resources to train[19]. The primary goal of the study to classify the banana plant leaves based on type of nutrients deficiency a particular leaf suffered from. Another goal is that solution should be less resource intensive and less image-specific inductive bias which means features are not restricted to focus on local regions they also include global patterns in making decisions. The study begins with discussion about existing models and problem definition with its purpose. The remaining part of the study as following sections related work shows studies utilize computer vision to solve the similar

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:51:39 UTC from IEEE Xplore. Restrictions apply.

problems. proposed section comprises detailed information about dataset, models and the implementation methodology, experimentation gives glimpse of implementation environment, sample data images and training curve, result section shows performance of proposed solution with sample outputs, and conclusion section conclude the study and share future direction for the study.

II. RELATED WORK

In past years researchers have proposed various studies on classification task on plants. These proposed studies operate on distinct datasets. Some of them are studied and discussed below. The importance of detecting malnutrition into leaves and uses convolution neural network for classification between healthy leaves and leaves based on the type of nutrient deficient in particular plant is highlighted in [1]. Reference [2] discussed various open opportunities for computer vision in managing and monitoring the stresses and diseases where they highlighted that some diseases are because of malnutrition in plant leaves if it was detected on early stage than it might be not reached to state where diseases starts affecting the production of the crop. Dataset of Indian state banana crops leaves images suffered from 8 different type of malnutrition and a class for healthy leaves, dataset opens new door for research in agriculture monitoring part which gives room to test and train diseases detection, classification leaves and predict survival for particular plant is discussed by [3]. Reference [4] worked with convolution neural network with fine tuning to classify the leaves in different classes based on deficiency of the nutrient and healthy status of the leaves, used the dataset of banana leaves having raw and augmented images releases from Indian banana tree fields. Reference [5] studied accuracy of the DCNN in nutrient deficient detection in rice crop images having 10 different classes based on type of malnutrition those plant leaves have, and fine tune the benchmarked DCNN models and perform experiments on different DCNN models. This experiment shows that nutrient deficiency detection is important step in monitoring and predicting crop yields.The applications of computer vision in agriculture sector to monitor crop nutrients using computer vision technology, study emphasis on tracking malnutrition of crop which in future may help to predict when fertilizers need to supplied into fields and based on nutrient what kind of fertilizer needs to be use is given by [6]. The role of the computer vision and machine learning in agriculture sector study which shows that computer vision alone plays vital role in monitoring of the crops in autonomous method, where detecting diseases on leaves, weed detection are areas in which computer vision is under testing and sooner will be available in real time systems is highlighted by [7]. A fine-tuned vision transform model for detecting infections and diseases over the plant leaves, and shows that vision transformer method which processes the images in patches surpasses the previous models in detecting diseases even having smaller regions affected through infection on plant leaves is proposed by [8]. A transformer leaf disease detection model which effective in terms of model size, even though transformer model in smaller size then also it outperforms over the existing CNN models in leaf regions detection task is given by [9].

III. PROPOSED SECTION

A. Dataset

Dataset was taken from Mendeley Data platform which was published on 24 oct 2022 by researchers from Malnad College of Engineering, Agriculture College, Tarekere, Hassan, Karnataka, India all the images were captured from high resolution mobile cameras from different locations which were Hassan , Ramnagara, Shimoga, Davangere and Hosapete districts of Karnataka, India. Original images dimensions are 3072 × 4096, 3096 × 4128, 30 0 0 × 40 0 0 but these all images are resized into 225 x 225 using adobe Photoshop as image processing tool. Dataset contains boron, calcium, potassium, iron, magnesium, manganese, sulphur, zinc classes which represents the deficiency of mentioned nutrient in each leaf. Natural background removed from all images and replaced with black background to avoid noises due capturing of other plants leafs and overlapping of background environment. Augmentations such as horizontal flip, vertical flip, and rotations were performed over the images to balance the number of samples for each class. Table 1 shows, number of samples of banana plant leafs based on type of deficiency of nutrient and in raw dataset number of image samples varies a lot which shows dataset was imbalanced, to make number around each other data augmentation such as horizontal and vertical flip, rotation, Gaussian blur and etc. were used. Dataset contains image samples for nutrient deficiency against boron, calcium, potassium, iron, zinc, manganese, magnesium and sulphur those counts were mentioned in table 1 against rows for raw and augmented samples. Total number of images in raw dataset were 3000 and after augmentation this count reaches to 8040.

TABLE I. SHOWS NUMBER OF IMAGE SAMPLES PER CLASS

Deficient nutrient	Number of image samples in raw form	Number of image samples after applying augmentation
Boron	100	900
Calcium	440	900
Potassium	86	900
Iron	160	900
Zinc	24	840
Manganese	210	900
Magnesium	730	900
Sulphur	400	900
Healthy	850	900
Total number of image samples	3000	8040

B. Pre-processing

Once difference between counts of images belongs to different malnutrion class looks balance, then multiple image preprocessing filters are operated over the dataset which highlights the pattern signifies the presence of particular type of nutrient deficiency. Images are taken under various lighting conditions to reduce the skew behavior of contrast histogram

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:51:39 UTC from IEEE Xplore. Restrictions apply.

equalization is applied over the greyscale version of the image which eliminates noises which creates false positives during classification. And then sharpening filter is applied over the images which highlight the areas effected with malnutrion in images.

C. Vision transformer

Core of the vision transformer is transformer based architecture for image handling tasks, where in order to process each part of the image it divides the images into patches then apply self-attention mechanism on each of these patches to get relationship between these different patches. Reducing dimensions of the patches is crucial part to handle, where vision transformer uses linear projections then leaned positions embedding's are added into encoder to encode the 2 dimensional topology of the image. Mathematically vision information transformer represented as $X \in \mathbb{R}^{H*W*C}$ which represents input image having height H, width W and channel C, where each such image divided into non-overlapping patches and each batch is linearized into a vector. Assuming P be a patches and D is the dimension of each patch after linearization.

$E \in \mathbb{R}^{P*D}$ is input embedding which is obtained by applying linear projection over each patch vectors.

$$E = X.W_{patch} + b_{patch} \quad (1)$$

Here in (1) b_{patch} is bais term and $W_{patch} \in \mathbb{R}^{P*D}$ is weighted matrix.In-order to incorporate spatial information E embeddings are added into positional embedding $PE \in \mathbb{R}^{P*D}$.

$$E_{pos} = E + PE \quad (2)$$

Then in (2) Epos is given as input into a transformer encoder in order to produce an output $Y \in \mathbb{R}^{P*D}$ .Y equals to TransformerEncoder(Epos) then ouput Y is classified using linear layer.

$$Z = Y.W_{class} + b_{class} \quad (3)$$

Here, $W_{class} \in \mathbb{R}^{D*num_classes}$ is weight matrix and b_{class} bias term. These equations represent the high-level mathematical model behind the working of vision information transformer, in proposed study ViT (vision information transformer) imported using transfer learning and then hyper parameters of baseline model are updated for specific use case of banana leaves classification based on malnutrition deficiency in leaf.

D. Dataset directory structure

Dataset divided into training, Validation and testing sets, directory structure for each split were created where training, validation and test splits samples were kept in folder names as per their nutrient deficiency. In final repositories each training, validation and test contains 9 folders where 8 against each nutrient deficiency and 1 for healthy leaf samples.

E. Transfer learning

Transfer learning is used to classification of image leafs based on deficiency of each nutrient in leafs. Here vit-base-patch16_224 pre-trained model is used on top of that model is fine tuned for leaf classification based on nutrient defiance task by adjusting model size, layer depth , optimization function and learning rate. In optimizing terms some layered are freeze as pertained model was trained on black color correctly in pre-processing stage backgrounds of all images were removed and replaced with black color so freezing some layer makes results better in terms of scores.

F. Hyper parameters tuning

Once pre-trained model imported parameters such as Patch size was tined for smaller size as a smaller patch size, like 16x16 able to capture the finer details in the leaf images, which makes model to identify complex patterns of diseases over leaf with better accuracy, moderate number of layers 10 was chosen for model which makes model to capture both low-level and high-level features from images even without making model computationally expensive. Feed forward dimensions set to 2048 as representation capacity needs to be balanced, 12 number of attention heads were used which capture various dependencies within the leaf images effectively, in order to generalize the model and avoiding over fitting drop rate set at 0.2 with all above defined values and values for weight decay was taken as 0.01, which helped in preventing over fitting and control the model's complexity during training. Values of all these final parameters were summarized in table 2.

TABLE II. SHOWS VALUES OF ALL FINAL PARAMETERS

Parameters	Values
Patch size	16X16
Number of layers	10
Feed-forward dimensions	2048
Attention heads	12
Drop rate	0.2
Weight decay	0.01

IV. EXPERIMENTATION

A. Implementation

Implementation of the proposed study begins with preprocessing the images which includes processes such as filtering noisy images, and resizing, after that images belongs to each class were count and if there was large gaps between the counting figures then augmentation was performed over the images, images augmented by rotations, translations , horizontals and verticals slicing after that complete dataset was divided into segments one for training and another for testing then arranging data into defined directory structure in which each training/testing had samples of ratios defined for each class, means difference between number of images of each class is lower for each training/testing sample set. Once dataset was prepared now, pertained model was loaded and parameters were finalized as defined in table 2 then start the training process over training set of data by initializing the model, then save the model file to perform testing over unseen data points which were in testing samples results of those were saved for evaluation purpose.

B. System

Experiment was performed over the system on windows 10 pro machine having 16 GB DDR4 Ram, and 512 GB SSD with

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:51:39 UTC from IEEE Xplore. Restrictions apply.

Nvidia GPU of 2 GB. On python version 8.5 with required packages such as numpy, pytroch, pandas, os, opencv, and tensorboard.

C. Data samples with labels

Images of each type of nutrient deficiency was coded in order to simplify the class labels, fig. 1 shows assigned labels for samples belongs to each class, with their labels such as boron represented by label 0, Calcium as label 1, Healthy as label 2, Iron as label 3 Magnesium as Label 4, Manganese as Label 5, Potassium as Label 6, Sulphur as Label 7 and Zinc as Label 8.

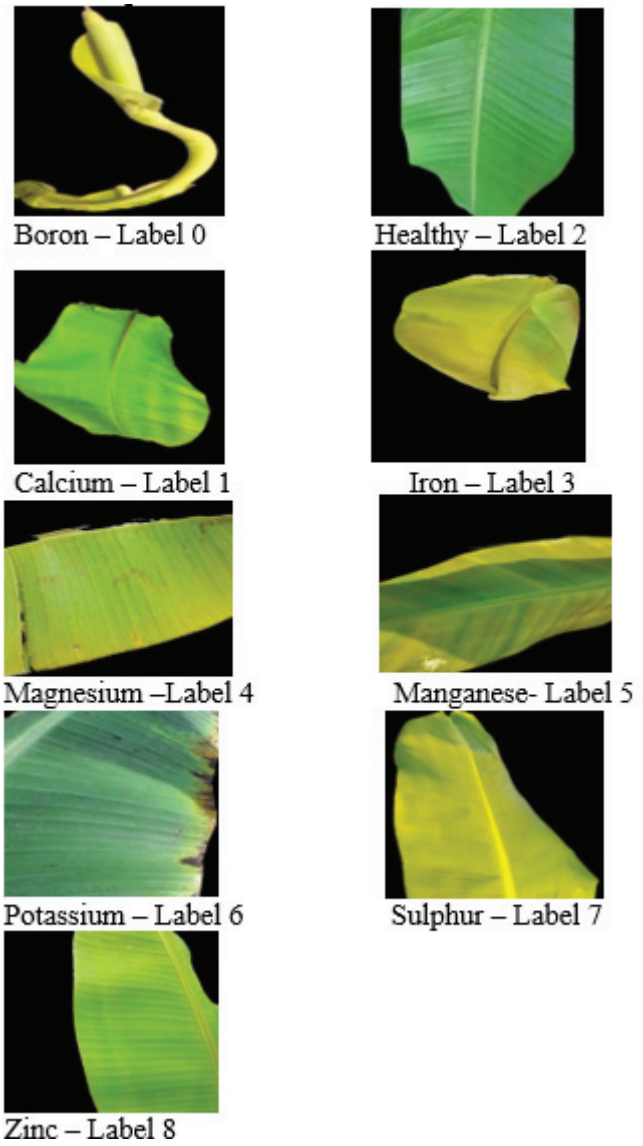


Fig. 1. Shows sample data images and their respective label classes

Images are pre-processed and augmented in order to balance the class labels so that model was not produce results which are biased for any particular class, and each sample's background was removed and replaced with black color background in-order to remove background noise, color variations and textures which makes model less complex and generic in nature.

D. Training loss curve

Once images were pre-processed, images were passed into initialized model for training. During training performance of the model was monitored through the logs generated corresponding to each epochs, here fig. 2 shows loss curve generated from logs saved during training in tensor scalar board which shows that reduction in training loss with respect to each epoch, it is observable from the curve that loss was reduced at every iteration and reached minimum and stable value at epoch 278, which makes to stop the training stage as there is no further updates in loss occur after iteration number 278, where loss seems to 0.06, and makes model stable. In curve vertical axis represented loss figures and horizontal axis represented the number of epochs.

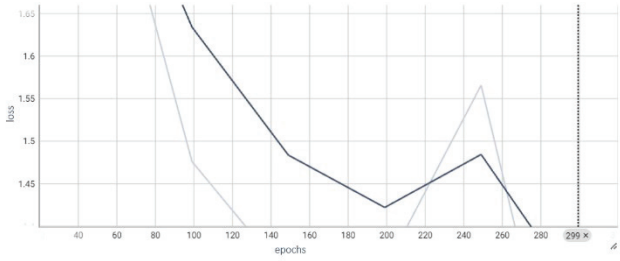


Fig. 2. Shows loss curves generated from logs saved during training in tensor scalar board

V. RESULTS

Once the model was trained testing set image predicted over the saved model file and performance was evaluated here table 3 shows proposed model performance scores for various classes which are different types of malnutrition in particular banana plant leaf, it is observable from the table that average accuracy for classes is around 90, for both training and testing sets of data which shows that model is generalized in nature and able to manage trade of between bias and variance.

TABLE III. SHOWS ACCURACY, PRECISION AND RECALL SCORES FOR EACH CLASS LABEL

Metrics / Class label	Training			Testing		
	Accuracy	*Precision*	*Recall*	*Accuracy*	*Precision*	*Recall*
Label 0	93	90	91	91	89	90
Label 1	91	89	90	90	87	89
Label 2	92	90	91	90	89	91
Label 3	92	90	89	92	90	89
Label 4	90	88	89	88	87	89
Label 5	91	89	87	87	90	88
Label 6	90	91	89	90	90	90
Label 7	91	89	90	89	88	89
Label 8	87	88	87	89	87	90

A. Illustrative examples

Fig. 3 shows the examples of output produced by the proposed model with their ground truth class and predicted label over some of the randomly selected image sample from testing set of the dataset, through illustrative examples it is observable that model performs quiet well even the images of the leafs looks similar such as in b part of figure 3 where samples of label 1 and label 2 looks similar but model able to classify.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:51:39 UTC from IEEE Xplore. Restrictions apply.

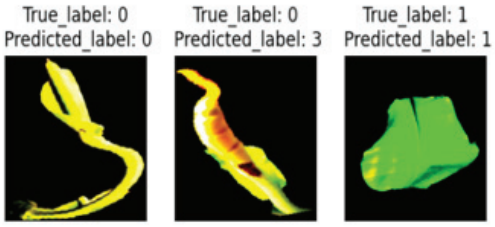


Fig. 3a Samples images of plants lacking boron and calcium

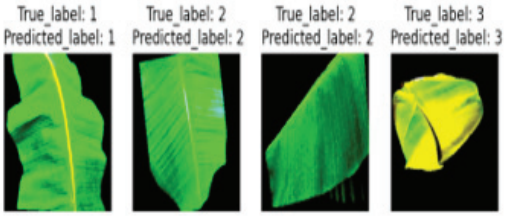


Fig. 3b Samples of healthy plant images, calcium and iron lacking images

Fig. 3. Shows sample output images and their respective label for both cases ground truth and predicted classes

Once results were evaluated for both training and testing set of images then scores of the convolutional neural network based existing solution[1] compared with proposed one here, table 4 compares the performance of convolutional neural network (CNN) [1] based classification model for classifying banana plant leaves based on nutrient deficiency with the vision information transformer based model, which surpass over the CNN model [1] in terms of recall and precision values, even with lower computational power, which shows that vision transformer are potential solutions for making decision over image taken from complex environments.

TABLE IV. SHOWS COMPARISON OF TEST RESULTS WITH CNN BASED SOLUTION

Metrics / Class label	CNN[1]		Proposed ViT	
	Precision	*Recall*	*Precision*	*Recall*
Label 0	79.73	97.4	89	90
Label 1	79.84	71.13	87	89
Label 2	85.32	94.89	89	91
Label 3	89.96	92.02	90	89
Label 4	73.95	81.09	87	89
Label 5	92.94	97.93	90	88
Label 6	96.74	96.88	90	90
Label 7	92.74	77.57	88	89
Label 8	87.57	63.79	87	90

TABLE V. SHOWS COMPARISON OF TEST RESULTS WITTH EXISTING SIMILAR STUDIES

Models	Accuracy	Recall	Precision
CNN[1]	80.1	85.85	86.53
ShuffleNet [4]	82.1	80	78.8
Inceptionv3[5]	79.84	80.6	87
ResNet50[5]	85.32	88.89	89
EfficientNet[18]	89.96	92.02	86
Proposed ViT	91	90	87

Table 5 represents comparison of the evaluation metrics between existing classification models based on deep convolutional neural networks with the proposed vision information transformer (ViT) based solution, and table 5 shows that creation of patches of images for understanding global pattern strategy of ViT makes it performs well over existing solution in terms of accuracy, precision and recall with handling tradeoff between precision and recall for similar task along with lower computational resources.

VI. CONCLUSION AND FUTURE SCOPE

Proposed study performed well for classification of bananas leaf's based on deficiency of multiple nutrients in the leaf's which ultimately helps in identification of reasons behind slow rate of improve in phenotype growth of plant and fruits with diseases. Proposed study uses transfer learning and pertained transformers to classify leaves, and generate models which were compact in size of existing solutions and better accuracy scores both at same time makes effective solution for mobile like portable devices applications even with lower hardware configuration.

REFERENCES

[1] K. A. M. Han, N. Maneerat, K. Sepsirisuk and K. Hamamoto, "Banana Plant Nutrient Deficiencies Identification using Deep Learning," 9th International Conference on Engineering, Applied Sciences, and Technology (ICEAST), Vientiane, Laos, 2023,pp. 5-9.

[2] Shin, J.; et. al.,"Trends and Prospect of Machine Vision Technology for Stresses and Diseases Detection in Precision Agriculture". AgriEngineering , vol 5, pp 20-39, May 2023.

[3] Sunitha P, Uma B, Channakeshava S, Suresh Babu C S,"A fully labelled image dataset of banana leaves deficient in nutrients" ,DatainBrief,vol. 48,p. 109155,june 2023.

[4] R. Guerrero, B. Renteros, R. Castañeda, A. Villanueva and I. Belupú, "Detection of nutrient deficiencies in banana plants using deep learning," 2021 IEEE International Conference on Automation/XXIV Congress of the Chilean Association of Automatic Control (ICA-ACCA), Valparaíso, Chile,2021, pp. 1-7.

[5] Zhe Xu, Xi Guo, Anfan Zhu, Xiaolin He, Xiaomin Zhao, Yi Han, Roshan Subedi, "Using Deep Convolutional Neural Networks for Image-Based Diagnosis of Nutrient Deficiencies in Rice", Computational Intelligence and Neuroscience, vol. 20, pp 1-12 pages, 2020.

[6] Peng, Q., Xu, W. "Crop Nutrition and Computer Vision Technology", Wireless Pers Commun, vol. 117, pp. 887–899 ,2021.

[7] Bansal, J.C., Uddin, M.S.,"Computer Vision and Machine Learning in Agriculture:, vol. 3, Springer, Singapore,2022,pp.15-34.

[8] Parez, S.; Dilshad, N.; Alghamdi, N.S.; Alanazi, T.M.; Lee, J.W." Visual Intelligence in Precision Agriculture: Exploring Plant Disease Detection via Efficient Vision Transformers", Sensors, vol. 15.pp. 15-23, 2023.

[9] Huy-Tan Thai, Kim-Hung Le, Ngan Luu-Thuy Nguyen, "FormerLeaf: An efficient vision transformer for Cassava Leaf Disease detection", Computers and Electronics in Agriculture,vol 204,pp. 1-13 2023.

[10] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in Proceedings of the IEEE, vol. 86, pp. 2278-2324, Nov. 1998.

[11] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E "ImageNet Classification with Deep Convolutional Neural Networks", Advances in Neural Information Processing Systems, vol 25,pp 84-90, 2012.

[12] Karen Simonyan, Andrew Zisserman "Very Deep Convolutional Networks for Large-Scale Image Recognition" ,2014 [Online], Avilable :arXiv:1409.1556.

[13] Christian Szegedy, et. al." Going Deeper with Convolutions " , 2014 [Online], Avilable :arXiv:1409.4842.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:51:39 UTC from IEEE Xplore. Restrictions apply.

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun ”Deep Residual Learning for Image Recognition”, 2015, [Online], Avilable : arXiv:1512.03385

[15] Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger “Densely Connected Convolutional Networks”, 2018, [Online], Avilable : arXiv:1608.06993

[16] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, Quoc V. Le “Learning Transferable Architectures for Scalable Image Recognition”,2018, [Online], Avilable : arXiv:1707.07012

[17] Andrew G. Howard, et. al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications” 2018, [Online], Avilable : arXiv:1704.04861

[18] Mingxing Tan, Quoc V. Le EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 2017, [Online], Avilable : arXiv:1905.11946

[19] Alexey Dosovitskiy, et. al. “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale” 2021, [Online], Avilable : arXiv:2010.11929.

[20] Olga Russakovsky, et. al.”ImageNet Large Scale Visual Recognition Challenge” 2015, [Online], Avilable : arXiv:1409.0575.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:51:39 UTC from IEEE Xplore. Restrictions apply.

2024 5th International Conference for Emerging Technology (INCET)
Karnataka, India. May 24-26, 2024

Comparative Analysis of various Slot in Patch Antennas: A Review

Sumit Kumar Khandelwal
Electronics and Communication Engg.
Delhi Technological University
Delhi, India
sumit.khandelwal@dtu.ac.in

Anupam Verma
Electronics and Communication Engg.
Delhi Technological University
Delhi, India
iamanupamv@gmail.com

Amarnath Yadav
Electronics and Communication Engg.
Delhi Technological University
Delhi, India
amarnathyadav312696@gmail.com

Arpan Goyal
Electronics and Communication Engg.
Delhi Technological University
Delhi, India
arpangoyal2002@gmail.com

***Abstract*—In the ever-evolving landscape of wireless communication, the demand for compact, high-performance antennas spanning a broad frequency range has become paramount. Microstrip patch antennas have emerged as a promising solution for fulfilling this demand, offering versatility and adaptability in various configurations. This research delves into a comparative analysis of different microstrip patch antenna designs tailored for broadband applications in the 2.4-6 GHz frequency range. As wireless technologies continue to advance, understanding the nuances of antenna performance becomes pivotal in achieving optimal communication systems. By scrutinizing the characteristics of E-shaped, U-shaped, Hexagonal, Circular, H-slot, Octagonal, and Rectangular microstrip patch antennas, this study aims to provide valuable insights that will inform the selection and design process for future wireless communication applications. The investigation spans crucial parameters such as bandwidth, gain, radiation pattern, return loss, impedance matching, and antenna efficiency, aiming to distill practical considerations for engineers and researchers working on the forefront of electronics and communication engineering.**

Keywords—Microstrip Patch Antenna, Gain, Return Loss, Impedance Matching, Antenna Efficiency, Bandwidth, U-shaped, Rectangular, E-shaped, Hexagonal, Circular, H-slot, Octagonal Microstrip Antenna.

I. INTRODUCTION

The microstrip antenna, a pivotal innovation in the domain of radio frequency engineering, has played a transformative role in the evolution of wireless communication systems. Its inception can be traced back to the mid-20th century when researchers began exploring planar transmission line technologies for antenna applications. The microstrip antenna, as we recognize it today, emerged as a result of continuous refinement and adaptation to meet the demands of modern communication.

The roots of the microstrip antenna can be found in the development of microstrip transmission lines in the 1950s. Initially conceived for microwave integrated circuits, researchers soon recognized the potential of these planar structures for radiating electromagnetic waves. The transition from transmission lines to antennas marked a significant shift, with notable contributions from innovators like Robert E. Munson and George F. Engen in the 1970s, who played key roles in refining the microstrip antenna design.

The fundamental structure of a microstrip antenna typically consists of a metallic radiating patch positioned on one side of a dielectric substrate, separated by an air or dielectric gap from a conductive ground plane on the other side. This planar configuration not only simplifies manufacturing processes but also lends itself to compact and lightweight designs.

II. LITERATURE SURVEY

The continual reduction in the size of mobile devices has spurred a significant surge in the development of compact antenna structures, supplanting conventional counterparts in mobile communication. Microstrip patch antennas have emerged as a highly promising solution for handheld electronic devices, thanks to their multiband capabilities and compact configurations. Comprising a rectangular patch on a dielectric substrate over a ground plane, these antennas can take diverse shapes, including octagonal configurations. Commonly fabricated with conducting materials such as Copper or Gold, microstrip patch antennas offer advantages like lightweight construction, cost-effectiveness, and simplicity in fabrication. However, despite their numerous merits, these antennas confront a substantial challenge in terms of low gain. [6] Leveraging the octagonal shape in microstrip patch antenna design optimally addresses the challenge of low gain, capitalizing on its compact and versatile configuration. This strategic utilization enhances antenna performance, offering a refined solution for efficient signal transmission in the evolving landscape of compact electronic devices.

A Microstrip patch antenna for 5G WLAN applications is proposed, operating at a frequency of 4.8 GHz and featuring an H-shaped slot within its rectangular patch. The incorporation of the H-shaped slot serves to enhance key antenna parameters such as VSWR, reflection coefficient, gain, and bandwidth.[4] The design is meticulously crafted to adhere to industrial standards set for 5G WLAN applications, ensuring optimal performance within this advanced communication context. The thin planar profile of the Microstrip patch antenna adds to its appeal, allowing for easy integration onto the flat surfaces of space-borne platforms such as aircraft and missiles. Notably, the antenna design yields a reflection coefficient of -25.44 dB and a VSWR of 1.11 at the specified frequency of 4.8 GHz, demonstrating its

2024 5th International Conference for Emerging Technology (INCET) | 979-8-3503-6115-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/INCET61516.2024.10593030

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:50:53 UTC from IEEE Xplore. Restrictions apply.

effectiveness in meeting stringent performance criteria for 5G WLAN applications.

In biomedical applications, the imperative for low-latency, high-bandwidth transmission of patient information necessitates advanced antenna solutions. This paper proposes a multi-band Hexagonal Shaped Meandering Antenna (HSMP) with a defected ground structure designed to enhance the bandwidth, return loss, and gain characteristics of a Microstrip Patch (MSP) antenna. Leveraging a slotted meandering approach for optimization, our experimental analysis reveals that the suggested HSMP technique outperforms other MSP antennas, exhibiting greater bandwidth and consistently higher gain.[2] Additionally, we advocate for the use of a single-layer substrate to craft a compact, shallow-rejection, low-profile microstrip patch antenna tailored for high-selectivity differential filtering. This research aims to contribute to the evolution of efficient data transmission in biomedical contexts.

The E-shaped microstrip antenna emerges as a distinctive and innovative variant in the realm of microstrip antenna designs, featuring a unique radiating element shaped like the letter 'E.' This configuration departs from traditional patch antenna shapes, providing engineers with an additional degree of freedom to tailor antenna characteristics.[3] The intricacies of the E-shaped microstrip antenna's design significantly influence key parameters such as resonance frequency, bandwidth, and radiation pattern, making it a compelling subject for in-depth exploration. As we delve into the analysis of this specific microstrip antenna type, the goal is to unravel the nuanced performance aspects that contribute to its suitability for various wireless communication applications. Understanding both the strengths and limitations of the E-shaped microstrip antenna is pivotal for engineers aiming to optimize antenna performance in the ever-evolving landscape of radio frequency engineering.

The U-shaped microstrip antenna stands out as an innovative configuration in microstrip antenna designs, featuring a distinctive radiating element shaped like the letter 'U.' This departure from traditional patch antenna shapes offers engineers a novel approach to tailor antenna performance for specific applications. The U-shaped design's structural intricacies play a crucial role in determining key parameters such as resonance frequency, bandwidth, and radiation pattern.[1] In exploring this microstrip antenna variant, the goal is to uncover nuanced performance characteristics that define its utility in modern communication needs. Notably, the U-shaped microstrip antenna presents advantages such as enhanced bandwidth, addressing wideband performance requirements. Its compact and space-efficient nature makes it ideal for integration into portable devices or systems with limited space. Additionally, the U-shaped geometry influences radiation characteristics, leading to improved directivity and radiation efficiency, further enhancing its appeal in wireless communication applications.

The circular microstrip antenna emerges as a notable and versatile design within the microstrip antenna landscape, characterized by its circular radiating element. This configuration has garnered attention for its unique advantages in wireless communication systems, particularly its ability to achieve an omnidirectional radiation pattern, making it well-suited for applications requiring consistent signal coverage in all directions. Delving into the intricacies of the circular microstrip antenna, we explore its design principles, fabrication simplicity, and performance attributes. Notably, the circular shape not only streamlines the fabrication process by eliminating complex structures but also contributes to cost reduction. Beyond its simplicity, these antennas often exhibit robust impedance matching, facilitating efficient power transfer between the antenna and the transmission line.[5] This multifaceted appeal positions circular microstrip antennas as valuable components in diverse communication scenarios, especially where signal uniformity, ease of fabrication, and impedance matching are paramount.

III. PARAMETERS

A. Gain

The review paper explores various antenna designs, each tailored to optimize specific parameters such as gain, return loss, and directivity. The hexagon-shaped antenna is crafted for high throughput biomedical applications, operating at 2.4 GHz with a VSWR less than 2 and achieving a gain of 3 dB[2]. This superior gain can be attributed to the hexagon's geometric characteristics, providing an increased aperture size that captures and concentrates more electromagnetic energy. The enhanced directivity, a result of the hexagon's complexity, further amplifies the gain. The optimization of return loss characteristics and the incorporation of sophisticated matching networks contribute to the overall efficiency of power transfer, making the hexagon-shaped antenna well-suited for higher gain applications.

Similarly, the circular microstrip patch antenna designed for 5G applications at 3 GHz stands out with a high gain of 10.0 dB[5]. Circular patch antennas achieve this high gain through their circular geometry, ensuring a uniform radiation pattern for omni-directional coverage. Symmetry in design facilitates focused and directive radiation, while resonance at the desired frequency ensures efficient energy radiation. Low cross-polarization and low side lobes contribute to the concentration of energy in the main lobe. Surface wave suppression minimizes energy loss, and the simplicity of their structure makes circular patch antennas cost-effective and easy to integrate.

In contrast, the U-shaped antenna excels in gain with a transmitter gain value of 8.019 dBi at the resonating frequency band of 2 to 4 GHz[1]. The U-shaped structure creates a resonant cavity that enhances radiation efficiency, allowing for the concentration and redirection of electromagnetic waves. This results in a more focused radiation pattern, promoting impedance matching, and maximizing power transfer. The curved arms of the U-shaped antenna act as directors, further concentrating radiation in the desired direction. This overall geometry allows for fine-tuning, optimizing performance for specific frequency bands, making the U-shaped antenna a robust choice for applications where high gain and directional radiation are crucial.

The E-shaped antenna, operating at about 6.23 dB gain over its frequency range, showcases its ability to achieve high gain attributed to several key factors[3]. The E-shaped configuration creates multiple radiating arms, increasing the effective aperture of the antenna. This larger aperture allows for the capture and concentration of more electromagnetic energy, resulting in improved gain. The arms of the E-shaped antenna act as directors, aiding in focusing radiation in a specific direction. The geometry allows better control and manipulation of the electric field, contributing to enhanced

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:50:53 UTC from IEEE Xplore. Restrictions apply.

radiation efficiency. Additionally, the design facilitates impedance matching, minimizing signal reflections, and optimizing power transfer, making the E-shaped antenna versatile and effective for applications requiring directional and efficient signal propagation.

Examining the octagonal patch antenna, the array reveals a gain increase of 6.25 dB compared to the single patch on the FR4 substrate[6]. The choice of substrate material significantly impacts the antenna's resonance frequencies and gain, with FR4 exhibiting better performance. The octagonal patch antenna demonstrates the importance of substrate selection and geometry in achieving higher gain, as well as the potential for array configurations to further enhance performance.

Lastly, the 4x1 array of rectangular patch antennas radiates with higher gain in the forward direction compared to the single patch. The array achieves a gain of 5.284 dBi, surpassing the 1.369 dBi gain of the single patch. The directivity of the array, measured at 10.59 dBi, outperforms the 6.759 dBi directivity of the single patch. This underscores the advantage of array configurations in achieving higher gain and directivity, crucial for specific applications.

In summary, each antenna design demonstrates its unique attributes in optimizing gain and related parameters. The hexagon-shaped antenna leverages its geometric characteristics, the circular patch antenna excels in uniform radiation, the U-shaped antenna emphasizes resonance and directional focus, the E-shaped antenna utilizes multiple radiating arms, the octagonal patch antenna emphasizes substrate material, and the rectangular patch array showcases the advantage of array configurations. These findings collectively contribute to the understanding of antenna design considerations for various applications, highlighting the significance of geometry, resonance, substrate material, and array configurations in achieving higher-gain

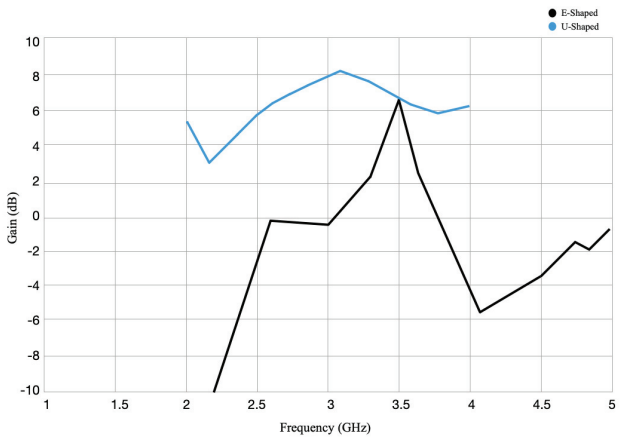


Fig. 1. E-Shaped vs U-Shaped Gain Comparison

B. Return Loss

Return loss is a measure of how much signal is reflected back to the source in a transmission system. It is commonly expressed in decibels (dB) and is defined as the ratio of the power of the reflected signal to the power of the incident signal, expressed as a negative dB value.

Return Loss (RL) can be calculated using the following formula:

$$RL\ (dB) = -20 * log10(\frac{Preflected}{Pincident})$$

•RL is the return loss in decibels (dB).

•$P_{reflected}$ is the power of the reflected signal.

•$P_{incident}$ is the power of the incident signal.

Return loss is an important parameter in the characterization of transmission systems, antennas, and other RF (Radio Frequency) components. It provides information about the amount of power that is not effectively transmitted through the system and is instead reflected back.

The U-slot antenna exhibits a remarkable -45.6 dB return loss, attributed to its innovative design that introduces an extra degree of freedom, facilitating enhanced impedance matching and a broader bandwidth[1]. Through simulations, the antenna demonstrates resonance at 3.145 GHz with an impressive bandwidth resolution of approximately 19.5%. Optimal dimensions, specifically an 18 mm length and 2 mm width for the U-slot, are identified, underscoring the significance of these parameters for optimal performance.

Comparatively, the hexagonal patch antenna achieves a return loss of -24.4 dB[2], although the paper doesn't explicitly detail the reasons. Presumably, the hexagonal shape and meandering design, coupled with a defected ground structure and slotted meandering approach, contribute synergistically to improved performance. Meanwhile, the octagonal counterpart exhibits a slightly lower -23.9 dB return loss[6], suggesting a close performance to the hexagonal shape, potentially hinting at the influence of shape geometry on the antenna's characteristics. Overall, the study underscores the importance of geometric variations in achieving superior antenna performance.

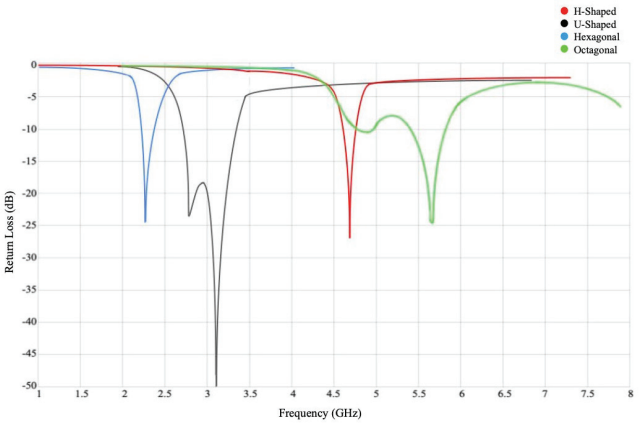


Fig. 2. H vs U vs Hexagonal vs Octagonal Shaped return loss Comparison

C. Impedance Matching

Impedance matching is a crucial concept in electronics, particularly in antenna design, to ensure efficient power transfer between components. It involves adjusting the electrical impedance of connected elements to achieve maximum energy transfer and minimize signal reflections. In the context of antennas, proper impedance matching enhances performance, optimizing signal strength and minimizing losses. Achieving impedance matching involves tuning parameters such as the dimensions of antenna elements and transmission lines. This ensures that the source

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:50:53 UTC from IEEE Xplore. Restrictions apply.

and load impedances align, preventing signal degradation and enhancing overall system efficiency, a vital consideration for reliable communication systems and electronic devices.

In antenna design, achieving optimal impedance matching is pivotal, and the U-slot antenna presented in this paper exemplifies a successful approach. Operating at 3.1245 GHz, it attains a notable impedance matching bandwidth of 19.5%, a result of meticulous adjustments to the U-slot and patch dimensions[1]. The prescribed slot length and width of 18 mm and 2 mm, respectively, emerge as optimal values for heightened performance, underscoring the critical role of these parameters in impedance optimization. Notably, the designed patch impedance aligns precisely with the resonant frequency, further contributing to effective impedance matching. The reported VSWR of 1.01 at the resonant frequency of 3.1245 GHz adds quantitative support to the antenna's impedance-matching prowess.

In contrast, the E-shaped antenna in the same study focuses on impedance matching between the radiating patch and the transmission feed line[3]. The authors highlight successful impedance matching to a 50Ω standard within the frequency range of the WiMAX operational band, ensuring good efficiency assets for the lanthanides-doped Magnesium ferrite antenna. This emphasizes the versatility of approaches employed in achieving impedance matching across different antenna designs.

D. Antenna Efficiency

Antenna efficiency is a critical metric that gauges the effectiveness of an antenna in converting electrical power into radiated electromagnetic waves. Expressed as a percentage, antenna efficiency accounts for various losses encountered in the antenna system, such as ohmic losses in the conductive elements, dielectric losses in the surrounding materials, and mismatch losses due to impedance mismatches. A highly efficient antenna optimally converts electrical signals into radiation, ensuring that a significant portion of the input power contributes to the desired electromagnetic field. Efficient antennas are pivotal in communication systems, as they enhance signal transmission range, improve overall system performance, and contribute to minimizing power consumption. Design considerations, material selection, and impedance matching are key factors influencing antenna efficiency, making it a crucial parameter in the development of reliable and high-performance wireless communication systems.

The review paper on the E-shaped antenna reports an efficiency of 73.32%[3]. The E-shaped antenna likely achieves this relatively high efficiency due to its carefully designed geometry. The distinctive shape of the E antenna is known for promoting desirable radiation characteristics, including improved directivity and radiation pattern control. The geometric configuration of the E-shaped antenna allows for better impedance matching, reduced losses, and enhanced radiation efficiency compared to some other antenna shapes. On the other hand, the H-slot antenna, with its increased directivity, experiences a boost in efficiency. This is because higher directivity often leads to a more focused and efficient use of radiated power[4]. In contrast, circular microstrip antennas may exhibit lower efficiency[5], possibly owing to broader radiation patterns and impedance matching challenges associated with their circular geometry. These differences highlight the importance of antenna design considerations in achieving optimal performance for specific applications, taking into account factors such as geometry, directivity, and impedance matching.

E. Bandwidth

The bandwidth of an antenna is a fundamental parameter that defines its ability to operate effectively across a range of frequencies. It represents the frequency span within which the antenna can maintain satisfactory performance, typically characterized by parameters such as acceptable radiation patterns, impedance matching, and gain. A wider bandwidth is desirable in communication systems as it allows the antenna to accommodate a broader range of frequencies, facilitating versatility in signal transmission and reception. Bandwidth considerations are particularly crucial in the context of modern wireless communication standards, where diverse frequency bands are allocated for different services. Antennas with broader bandwidths are better equipped to adapt to dynamic and evolving communication environments, making them essential components in systems requiring flexibility and adaptability to various frequency allocations and standards. Achieving optimal bandwidth often involves careful design considerations, including antenna geometry, feeding techniques, and impedance matching strategies, ensuring the antenna's effectiveness across a diverse spectrum of frequencies.

The bandwidth analysis of various microstrip antenna configurations reveals intriguing insights into their performance characteristics. The U-slot microstrip antenna, examined through S-parameter curves at a 10dB threshold[1], demonstrates a bandwidth of approximately 450MHz. However, the proposed Hexagonal patch antenna (HSPA) outshines this with a measured bandwidth of 3.2GHz, constituting a substantial 64% impedance bandwidth[4]. This improvement underscores the efficacy of the HSPA design in achieving a significantly wider frequency range, aligning well with the demands of High-Speed Packet Access (HSPA) technologies. Similarly, the utilization of an E-shaped slot proves advantageous, exhibiting enhanced bandwidth compared to a conventional microstrip antenna.

Moreover, the integration of an H-shaped patch contributes to bandwidth augmentation, with a bandwidth of approximately 140MHz derived from the S-parameter vs frequency curve[4]. Moving beyond shape variations, substrate choices also exert a considerable impact on bandwidth. For octagonal microstrip antennas across different substrates, variations in resonance frequencies and corresponding bandwidths are evident. For instance, the FR4 substrate resonates at two frequencies, resulting in a bandwidth of 200MHz and 500MHz[6], showcasing its versatility. Duroid resonates once but boasts a substantial bandwidth of 1200MHz, highlighting its superior frequency range coverage. Arlon resonates once with a bandwidth of 500MHz, and Rogers resonates once with a bandwidth of 100MHz, further emphasizing the influence of substrate selection on antenna bandwidth. These findings collectively underscore the importance of thoughtful design and material choices in optimizing the bandwidth of microstrip antennas for diverse communication applications.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:50:53 UTC from IEEE Xplore. Restrictions apply.

TABLE I. COMPARISON OF DIFFERENT MICROSTRIP ANTENNA

	Operating Frequency (GHz)	Gain (dB)	Directivity (dB)	Return Loss (dB)	Bandwidth (GHz)
Rectangular[8]	2.4	-	5.63	-54.99	0.081
E-Shaped[3]	3.5	6.93	7.58	-9.79	0.14
U-Slot[1]	3.12	8.01	-	-45.85	0.45
Hexagonal[2]	2.4	3.01	6.2	-24.4	3.2
Circular[5]	3	10.7	6.623	-6.689	-
H-Slot[4]	4.8	-	6.46	-10.7	0.14
Octagonal[6]	4.7-4.9	3.6	-	-10.5	0.2
Octagonal[6]	5.3-5.8	3.6	-	-23.9	0.5

IV. CONCLUSION

The review paper delves into various antenna designs, each tailored to optimize specific parameters such as gain, return loss, and directivity. The hexagon-shaped antenna emerges as a frontrunner for high throughput biomedical applications, operating at 2.4 GHz with a VSWR less than 2 and achieving a gain of 3 dB. This superior gain stems from the hexagon's geometric characteristics, providing an increased aperture size that captures and concentrates more electromagnetic energy. The enhanced directivity, a result of the hexagon's complexity, further amplifies the gain.

The circular microstrip patch antenna, designed for 5G applications at 3 GHz, stands out with a high gain of 10.0 dB. Circular patch antennas achieve this high gain through their circular geometry, ensuring a uniform radiation pattern for omni-directional coverage. Symmetry in design facilitates focused and directive radiation, while resonance at the desired frequency ensures efficient energy radiation. Low cross-polarization and low side lobes contribute to the concentration of energy in the main lobe. Surface wave suppression minimizes energy loss, and the simplicity of their structure makes circular patch antennas cost-effective and easy to integrate.

In contrast, the U-shaped antenna excels in gain with a transmitter gain value of 8.019 dBi at the resonating frequency band of 2 to 4 GHz. The U-shaped structure creates a resonant cavity that enhances radiation efficiency, allowing for the concentration and redirection of electromagnetic waves. This results in a more focused radiation pattern, promoting impedance matching, and maximizing power transfer. The curved arms of the U-shaped antenna act as directors, further concentrating radiation in the desired direction. This overall geometry allows for fine-tuning, optimizing performance for specific frequency bands, making the U-shaped antenna a robust choice for applications where high gain and directional radiation are crucial.

REFERENCES

[1] S. K. Dargar, A. Gupta, M. Sabir and A. Lakshmi, "Design of U-Slot Microstrip Patch Antenna for Wireless Applications," 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2022, pp. 1-5, doi: 10.1109/ICAECT54875.2022.9807838.

[2] S. Elangovan., R. Gokul Raj., M. Vethaprasath. and K. Mahendran, "Design and analyzing of hexagon-shaped microstrip patch antenna for biomedical applications," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-5, doi: 10.1109/ICCCI56745.2023.10128199.

[3] Alagarsamy, Sindhuja & Naidu, Vasant & Vinoth, P & Pandimadevi, M. (2012). Design and Evaluation of E-Shaped Microstrip Patch Antenna (MPA) For Wimax Application. International Journal of Computer Applications. 45. 34-37. 10.5120/6856-9449.

[4] A. Madankar, V. Chakole and S. Khade, "H-slot Microstrip Patch Antenna for 5G WLAN Application," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 1142-1144, doi: 10.1109/ICISS49785.2020.9315914.

[5] S. P. Pavithra and K. Vidhya, "Design of a Circular Microstrip Patch Antenna with High Gain for 5G Applications using 5G Millimeter Wave Bands Comparing to Rectangular Microstrip Patch Antenna," 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 30-34, doi: 10.1109/ICTACS56270.2022.9988232.

[6] Gabari, A.A., Yunusa, Z., Hamidon, M. N. (2019). Octagonal Microstrip Patch antenna array with gain enhancement for WiMax and WLAN Applications. Journal of Innovative Science and Engineering. 3(2): 27-39.

[7] Yadav, Ravindra & Kishor, Dr & Lal, Ram. (2011). Design of Hexagonal Patch Antenna for Mobile Wireless System. International Journal of Science, Technology and Management. Vol. 2. 34-38.

[8] H. Kaschel and C. Ahumada, "Design of Rectangular Microstrip Patch Antenna for 2.4 GHz applied a WBAN," 2018 IEEE International Conference on Automation/XXIII Congress of the Chilean Association of Automatic Control (ICA-ACCA), Concepcion, Chile, 2018, pp. 1-6, doi: 10.1109/ICA-ACCA.2018.8609703.

[9] Garg, Madhuri & Singh, Nisha & Verma, Anupam. (2015). Rectangular Microstrip Patch Antenna. 6. 253-257.

[10] Bansal, Aakash & Gupta, Richa. (2018). A review on microstrip patch antenna and feeding techniques. International Journal of Information Technology. 12. 10.1007/s41870-018-0121-4.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:50:53 UTC from IEEE Xplore. Restrictions apply.

2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) | 979-8-3503-7156-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/AMATHE61652.2024.10582108

Complementary Dynamics of Global and Local Centrality for Influence Maximization in Social Networks

Rajesh Kumar Yadav
Computer Science and Engineering
Delhi Technological University,DTU
Delhi, India
rkyadav@dtu.ac.in

Aaryan Arora
Computer Science and Engineering
Delhi Technological University,DTU
Delhi, India
aaryan2134@gmail.com

Aayush Patel
Computer Science and Engineering
Delhi Technological University,DTU
Delhi, India
apaayush2003@gmail.com

Aditya Gupta
Computer Science and Engineering
Delhi Technological University,DTU
Delhi, India
adigupta2k2@gmail.com

***Abstract*—With the growing interest and ease of communication, the number of people using various social networks to disseminate information daily has increased exponentially. Social networking platforms have experienced exponential growth in recent years. To maximize the reach of information, there is a need to effectively identify influential nodes in social networks. Still, existing centrality measures and other methods have drawbacks relating to generalizing across networks and providing consistently accurate identification. In this study, a novel method is proposed by using the Global Local Influence (GLI) approach in particular combining closeness centrality and eigenvector centrality respectively. This aims to get more accurate identification of influential nodes across various networks. To evaluate the performance of the proposed work, experiments were conducted on synthetic and real-world datasets. The results were evaluated with the help of the diffusion models, namely Susceptible-Infected-Recovered (SIR) model and Node Removal Procedure (NRP). The results showed that the proposed algorithm outperforms the existing centrality measures by over 9% across the datasets supported by both the evaluation metrics used.**

Keywords—Closeness Centrality, Eigenvector Centrality, Social Networks, Node Removal Procedure (NRP), Susceptible-Infected-Recovered (SIR) Model

I. INTRODUCTION

In recent times, social networking has become an increasingly important way to establish interpersonal connections and for sharing information. It plays a vital role in modern society by providing the seamless sharing of information. The social networking platforms like Facebook, X, Instagram and LinkedIn have experienced exponential growth in the last few years. Currently, worldwide there are overall 4.95 billion global users across more than 130 social media platforms and is expected to reach 5.85 billion by 2027 [1].

A social network can be defined as a collection of nodes (individuals, clubs, organizations, etc.) which are connected by edges/links that indicate the association between them. It has many applications in various areas such as organisational studies, psychology, economics, information science and criminology [1] thus providing new avenues for the application of social network analysis (SNA) methods.

SNA focuses on information exchange among individuals/groups/organizations, with applications in identifying influential nodes within complex systems like social networks. This method helps to track the spread of information by considering influential nodes/users in message dissemination, offering insights into the diffusion of rumours and potential interventions to control the spread of rumours [2].

The existing methodologies for identifying influential nodes incorporate measures such as degree, closeness, and betweenness centrality [3]. While degree centrality considers the number of connections of a node, it has drawbacks in terms of accurately assessing centrality, especially when nodes with higher connections are additionally compared to more central nodes with fewer connections [4]. Closeness centrality measures the distance between nodes, while betweenness centrality evaluates a node's role as a bridge between other nodes in the network. While these measures outperform degree centrality, their use in large complex networks is difficult because of their high computational complexity [3].

In k-shell algorithm [5], the outer layers of a network are stripped away, and the core nodes are considered to be more influential. However, in the case of some special networks, it has the drawback of assigning the same influence value to a large number of nodes.

The discussed approach above assumes a node's influence is only dependent on its importance. Information entropy [8] introduces the consideration of neighbours' importance in assessing a node's influence. Also, various random walk models like PageRank [5] and LeaderRank [6] are used in identifying influential nodes in directed networks but are unable to work effectively in undirected networks.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:53:46 UTC from IEEE Xplore. Restrictions apply.

While there have been many efforts to identify influential nodes in social networks, there is still room for improvement in terms of accuracy and efficiency. Further research is needed to address the limitations of existing methods and to develop new approaches to better understand the importance of nodes in real-world networks. This study explores the use of Global Local information [3] for identifying influential nodes and evaluates its performance through experiments on real-world and synthetic datasets using SIR and NRP evaluation metrics.

The Global Local Information approach represents a novel method for analysing complex networks, specifically for identifying influential nodes. It addresses limitations in existing key node identification methods that consider either local or global information. The proposed method incorporates both global and local influences as two main factors, the local influence of a node is calculated using eigen vector centrality and closeness centrality has been used to evaluate the global influence of a node over the network [7]. This provides more accurate node influence estimation compared to a single centrality measure/algorithm. Global information is utilized to calculate the overall influence of nodes, while local information is employed to calculate their influence and that of their adjacent nodes. Distinguishing nodes with the same global influence enhance the differentiation effect [14].

The key contributions of the proposed research are as follows:

- This paper introduces a comprehensive evaluation of the proposed algorithm for identifying influential nodes in social networks.
- The proposed algorithm will identify influential nodes with a substantial impact on network structure and properties when removed from the network.
- The exploration of influential nodes in complex networks, allows targeted interventions to leverage their influence.
- The GLI approach makes noteworthy contributions to the field of SNA, particularly in dealing with extensive and intricate networks by incorporating both global and local information.

The subsequent sections of this paper are organized as follows: Section II reviews previous work in this field, Section III elucidates preliminary concepts, Section IV details the proposed methodology, Section V discusses the analysis results, and finally, Section VI provides concluding remarks and outlines the scope for future work.

II. Related work

This section includes research done in past in the field of identifying influential nodes in SNA. Already work has been done to quantify the significance of nodes in complex networks in several studies.

Centrality measures like degree, betweenness, closeness, and eigen vector centrality [2] are used to gather information on influential nodes based on their position and connections in the network. The clustered local degree centrality identifies important nodes by evaluating the local clustering of a node and the degree of its neighboring nodes [9].

In the heuristic clustering method, it allocates nodes into clusters based on the similarity between each pair of nodes present in the network. The most influential node in each cluster is chosen as a seed member. Despite its low time complexity, it has low accuracy in terms of identifying the optimal set of influential nodes [10] due to its possibility of getting trapped in locally optimal solutions.

Prior methods fall short of capturing the complete dynamics, that's why random walk models are used. These measures are integrated through preference analysis in the random walk algorithm [6] to rank influential nodes in a network. Also, the PageRank algorithm to address influence maximization problems is proposed. Zareie [11] proposed the gray wolf optimization algorithm as an optimization problem-solving approach, that considers the influence of nodes along with the distance between them. This shows improved performance with lower computational time compared to earlier heuristic methods used in SNA.

Panfeng et al. [12] introduced the VoteRank++ algorithm, which iteratively selects influential nodes where nodes with different degrees are assigned different numbers of votes. It prioritizes the nodes in terms of their voting ability, allowing each node to vote differently for its neighboring nodes. The Explosion-Trust (ET) Game Model [13] emphasizes how degrees of rumor explosion and trust in the source node can build a rumor-spreading model with the help of game theory.

While each measure is reasonable from its perspective, a comprehensive evaluation cannot be achieved individually. Hence, multiple hybrid approaches should be considered for the influence maximization problem. Utilizing deep learning techniques [13] with network embedding and deep reinforcement learning has also been considered for improved results. Liqing [7] emphasizes global selection based on the local influence algorithm, highlighting its two-stage filtering strategy and submodular property for enhanced efficiency and accuracy.

Within the research community, one can see that considerable work has already been done to find the influential nodes in complex networks. Despite the advances, there are still some drawbacks to the current studies. For example, centrality measures can be unreliable in capturing the true influence of a node in a network because they are based on limited information and can be influenced by the structure of the network itself.

This limitation has been addressed by a recent addition to the literature, Global Local Information. In this study, the GLI algorithm is used to identify influential nodes in a network based on existing literature. To achieve a balanced trade-off between effectiveness and efficiency, this paper proposes a local-global influence indicator-based algorithm and compares its performance to traditional centrality measures using real-world and synthetic datasets.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:53:46 UTC from IEEE Xplore. Restrictions apply.

III. Preliminaries

This section provides foundational knowledge of the concepts necessary for a deeper comprehension of the proposed research.

A. Global and Local Influence

In network analysis, a node's influence is assessed on both global and local scales. This evaluation is crucial for understanding the node's significance in shaping network dynamics and behaviours.

Globally, a node's influence is determined by its impact on the overall network structure. This is quantified using centrality measures such as closeness, betweenness, and eigenvector. These metrics assess the node's importance in facilitating communication or information flow across the entire network. On a local scale, a node's influence is evaluated within its immediate neighbourhood or community using metrics such as degree centrality and PageRank. These measures help understand how a node contributes to the cohesion and dynamics of its local network surroundings.

Understanding both global and local influence is essential for gaining comprehensive insights into the role of nodes within a network, facilitating the analysis of network structure, dynamics, and the implications for various network-driven phenomena.

B. Information Diffusion Models

Information diffusion models are mathematical frameworks extensively employed to examine how information propagates within social networks. These models enable the study of the dynamics of information flow and the factors influencing the speed and scope of diffusion in social networks. To validate the proposed method's effectiveness, Node Removal Procedure (NRP) and Susceptible Infected Recovered (SIR) models have been utilised to evaluate the rationality of node ranking outcomes.

a. Susceptible Infected Recovered (SIR)

The SIR model is a predominant, deterministic diffusion model utilised in social network analysis, exploring the dissemination of infectious diseases, rumours, and information within networks. This model assesses the entire population susceptible to the disease, estimating the number of infected individuals over a given period to forecast disease progression. Individuals or nodes within the network are categorised into three distinct states:

- Susceptible (S(t)): Number of people who are susceptible at time *t*
- Infected (I(t)): Number of infected people who can transmit disease at time *t*
- Recovered (R(t)): Number of recovered people at time *t*

The set of differential equations that describe the dynamics of disease transmission and recovery in a network can be expressed as follows:

$$\frac{dS}{dt} = -\beta SI \tag{1}$$

$$\frac{dI}{dt} = \beta S I - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

The transmission coefficient, denoted by β, governs the infection rate, and γ signifies the recovery rate. It is assumed that recovered nodes maintain permanent immunity. Temporal variations in the susceptible (S), infected (I), and recovered (R) populations can be examined.

b. Node Removal Procedure (NRP)

Node Removal Procedure (NRP) is a method employed in social network analysis to evaluate the impact of removing specific nodes from a network. It involves systematically removing nodes based on certain criteria, such as centrality measures or other relevant indicators. By observing the changes in network dynamics after each node removal, insights can be made into the influence of nodes within the network.

IV. Proposed Methodology

A. Brief Description of Proposed Algorithm

Within the domain of social network analysis, many centrality algorithms play a pivotal role in explaining the influence of nodes within the network. Each algorithm has a distinct methodology and provides a singular perspective on the significance of nodes. These algorithms consider variables such as connectivity, proximity, and the structure of the network.

The proposed algorithm makes use of the centrality algorithms to generate a GLI score for the nodes in the network. The GLI score is derived by combining a global and a local centrality measure. For the study, closeness centrality is taken to account for global influence of the nodes and eigen vector centrality is considered for local influence of a node.

a. Closeness Centrality

Closeness centrality measures how close a vertex is to all other vertices in the graph [15]. It quantifies how quickly information can spread in a network, indicating the efficiency of communication pathways. It is computed by finding the average of the shortest path lengths from the node to all other nodes and then taking the reciprocal. Mathematically, for a graph of vertices, V, and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:53:46 UTC from IEEE Xplore. Restrictions apply.

edges, E, $G = (V, E)$, the closeness score of a vertex v, $C_s(v)$, can be defined as:

$$C_s(v) = \frac{V\text{-}1}{(\sum_{u_j \in V,\ u_j \neq v} d(vu_j))} \quad (4)$$

where d(vu) is the distance between vertices v and u.

Nodes with higher closeness centrality are considered more central, facilitating quicker information dissemination. This metric is valuable for assessing network resilience and efficiency across various contexts. This study utilises closeness centrality as a measure of global influence. The selection of closeness centrality was driven by two factors: (a) it indicates nodes as more central if they are closer to most of the nodes in the graph, corresponding to visual centrality [16], and (b) it considers both direct connections and the efficiency of communication pathways.

b. Eigen Vector Centrality

Eigenvector centrality quantifies a node's influence in a network, emphasizing the quality of connections over mere quantity. It posits that links to highly connected nodes enhance a node's centrality more significantly. It is computed by solving an eigenvector equation, involving the network's adjacency matrix, A; the centrality score matrix, x; and, the largest eigenvalue, λ. Formally,

$$Ax = \lambda x \quad (5)$$

The solution of (5) gives the eigenvector centrality for a node v, x_v.

This approach is adept at pinpointing key influencers in networks where connections vary in importance, such as in social or citation networks. By focusing on the principle that a node's significance is bolstered by its connections to other influential nodes, eigenvector centrality offers a nuanced perspective on node importance, balancing the insights into a node's network value beyond simple connectivity. Hence, eigenvector centrality is used to compute the local influence of a node in this study.

B. Algorithm Description

The suggested algorithm finds influential nodes within an intricate network by incorporating the concept of appraising the complementary impact of a node's global and local influence. The research objectives outlined in this report primarily pertain to undirected and unweighted networks, denoted as $G = (V, E)$. This section introduces the key definitions and the proposed algorithm which are specifically relevant to an undirected, unweighted network concerning a node (v_i).

- Global Influence, $GI(v_i)$: The global influence of a node can be calculated using the closeness centrality. Mathematically,

$$GI(v_i) = C_s(v_i) \quad (6)$$

- Local influence, $LI(v_i)$: The local influence of a node quantifies its influence on its immediate neighbourhood. In this research, this definition is extended to consider the influence of a node on its community. In a social network, communities exhibit strong internal connections while having weaker connections to nodes outside it. To calculate the community (or local) influence of a node, the eigenvector centralities of its immediate neighbours are considered. If Z is the set of neighbors of v_i,

$$LI(v_i) = Harmonic_Mean(x_u,\ u \in Z) \quad (7)$$

The choice of the harmonic mean is justified by its reduced sensitivity [17] to extreme values compared to other averages, such as the arithmetic mean. This property makes it particularly suitable for analysing networks so that extreme values do not bias the results.

The node influence $NI(v_i)$ depends on $LI(v_i)$ and $GI(v_i)$ and is calculated as:

$$NI(v_i) = LI(v_i) + GI(v_i) \quad (8)$$

Higher *NI* score is desirable. A node with higher *NI* score is considered more influential. Conversely, a node with lower *NI* score is considered less influential.

V. Results

This section lists the results analysis of the experiments conducted as part of the study. SIR model was used to simulate how well the top K% influential nodes identified by each algorithm were able to spread influence in the network. Initially, top 1% nodes were infected with the rest of the nodes considered as susceptible. The infection factor, β, was set to 0.15 and the recovery factor, γ, was set to 0.1. An average of 100 runs was considered to determine the results of the model.

NRP is used for evaluating the accuracy of identified influential nodes in a network. The algorithms considered for the study provide a ranked list of influential nodes. Nodes are removed in sequence from the network according to this ranking. At each step, two key metrics of the resulting graph are assessed: the size of the Giant Component (GC) and the number of Connected Components (CC). An algorithm is deemed superior if it has a higher number of connected components while having a lower size of the giant component.

A. SIR

Fig. 1 shows the results obtained with SIR simulations. The graphs shown are plotted: where x-axis representing epochs and y-axis representing the sum of infected and recovered nodes. The plotted graphs depict the spread of information in the network with epochs. A steeper slope in the graph denotes a faster information spread which in turn indicates more accurate identification of initial influential nodes.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:53:46 UTC from IEEE Xplore. Restrictions apply.

As shown in fig. 1, the proposed algorithm shows a slightly faster rate of information diffusion than the other algorithms, indicated by its steeper slope. Additionally, the initial spread is much faster showcasing the superiority of the proposed algorithm in quick information spread before saturation. It reaches the peak faster than the rest of the algorithms in all datasets.

B. Node Removal Procedure

In fig. 2, the graphs show the results obtained after performing the node removal procedure on the datasets as discussed previously. Area under the curve (AUC) for size of GC and number of CC is plotted to perform a comprehensive comparison between the algorithms.

The proposed algorithm outperforms all algorithms in fig. 2(a) and (b) and fig.2 (e) and (f) datasets using both AUC - size of GC and AUC – number of CC. It has the highest AUC for the number of CC and the lowest size of GC amongst all the algorithms in these datasets. Additionally, in fig.2 (c) and (d), it shows superior performance according to the number of CC and just slightly worse than closeness according to the size of GC. This result highlights the superior performance of the proposed algorithm in identifying the influence spreading ability of the nodes and ranking them.

One limitation of the proposed approach is also highlighted where due to its focus on considering both the local and global factors it may perform worse than individual algorithms in specific datasets. Though due to its use of both the factors, it has less variable performance across datasets and performs better by 9% on average from the algorithms compared.

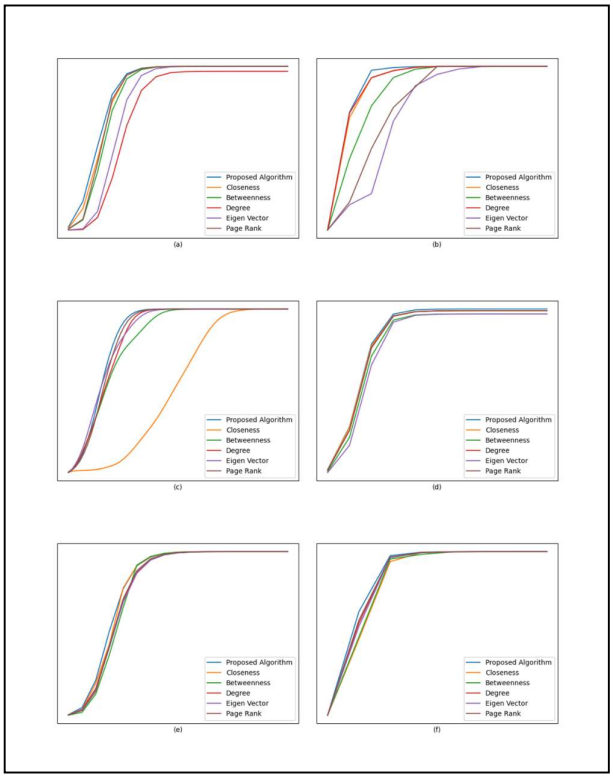


Fig.1: Results for NRP for various datasets

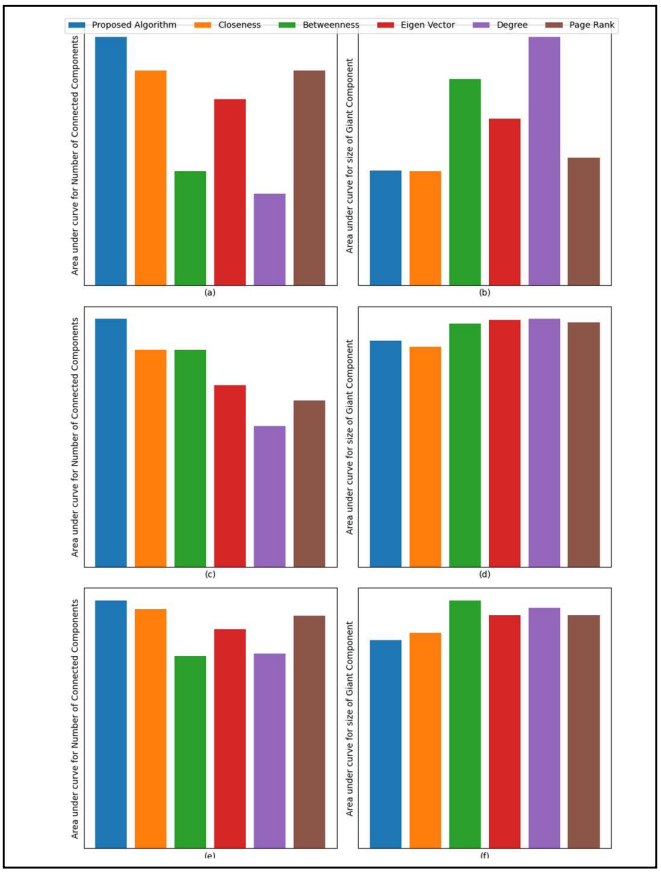


Fig.2: Results for SIR for various datasets

VI. Conclusion

SNA has shown increasing importance thanks to the era of online social networks having billions of daily active users engaging with multiple users. These networks have become very complex making it a challenge to effectively identify influential nodes. The role of influencers in the current society highlights their importance and in turn their power in the dissemination of information. In this study, a novel methodology is proposed using the GLI approach in particular combining closeness centrality and eigenvector centrality respectively.

The proposed algorithm was evaluated against multiple influential node identification algorithms using both real-world and synthetic datasets. The two diffusion models used for the comparison were SIR and NRP evaluation metrics. The results illustrated that the proposed algorithm performs better across all datasets supported by both the evaluation metrics used.

Though it does have its limitations where it was observed that it occasionally performs worse than its counterparts marginally. Additionally, as it requires the computation of both global and local measures namely closeness centrality and eigenvector centrality, it suffers from higher computation time than both of them.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:53:46 UTC from IEEE Xplore. Restrictions apply.

Further study in the field of influential nodes is required to find an algorithm that offers greater accuracy more efficiently and is a field of constant research as no single solution exists. The need to select the suitable algorithm for a specific use case is still required with not a single algorithm available to rank influential nodes effectively and efficiently in varying network structures, sizes and complexities.

REFERENCES

[1] A. Ficara, G. Fiumara, P. De Meo, A. Liotta, "Correlations Among Game of Thieves and Other Centrality Measures in Complex Networks," in G. Fortino, A. Liotta, R. Gravina, A. Longheu (eds), *Data Science and Internet of Things*, Internet of Things, Springer, Cham, 2021.

[2] B.A. Bhuiyan, "An overview of game theory and some applications," *Philos Progress*, vol. 59, no. 1–2, pp. 111–128, 2018.

[3] H. Hu, Z. Sun, F. Wang et al., "Exploring influential nodes using global and local information," *Sci Rep*, vol. 12, article 22506, 2022,

[4] J. Golbeck, "Analyzing networks," in *Introduction to Social Media Investigation*, Syngress, 2015, pp. 221-235, ISBN 9780128016565.

[5] J. Sheng, J. Zhu, Y. Wang, B. Wang, Z. Hou, "Identifying Influential Nodes of Complex Networks Based on Trust-Value," *School of Computer Science and Engineering, Central South University, Changsha 410083, China*.

[6] C. Mao, W. Xiao, "A Comprehensive Algorithm for Evaluating Node Influences in Social Networks Based on Preference Analysis and Random Walk," *Complexity*, 2018, pp. 1-16,

[7] L. Qiu, X. Tian, S. Sai, C. Gu, "LGIM: A Global Selection Algorithm Based on Local Influence for Influence Maximization in Social Networks," *IEEE Access*, vol. 8, pp. 4318-4328, 2020, doi: 10.1109/ACCESS.2019.2963100.

[8] S. Gao, J. Ma, Z. Chen, G. Wang, C. Xing, "Ranking the spreading ability of nodes in complex networks based on local structure," *Phys A*, vol. 403, pp. 130–147, 2014.

[9] M. Li, R. Zhang, R. Hu, F. Yang, Y. Yao, Y. Yuan, "Identifying and ranking influential spreaders in complex networks by combining a local-degree sum and the clustering coefficient."

[10] Y. Chen, X. Cao, K.J. Liu, "Community detection in networks: a game-theoretic framework," *EURASIP J. Adv. Signal Process.*, vol. 60, 2019.

[11] A. Zareie, A. Sheikhahmadi, M. Jalili, "Identification of influential users in social network using gray wolf optimization algorithm," *Expert Systems with Applications*, vol. 142, 2020, 112971, ISSN 0957-4174.

[12] P. Liu, L. Li, S. Fang, Y. Yao, "Identifying influential nodes in social networks: A voting approach," *Chaos, Solitons & Fractals*, vol. 152, 2021, 111309, ISSN 0960-0779.

[13] C. Guo, X. Chen, D. Chen, H. Gao, L. Yang, J. Ma, "Influential nodes identification in complex networks via information entropy," *Entropy*, vol. 22, p. 242, 2020.

[14] M. Mukhtar, Z. Abal Abas, A.S. Baharuddin et al., "Integrating local and global information to identify influential nodes in complex networks," *Sci Rep*, vol. 13, article 11411, 2023.

[15] K. Okamoto, W. Chen, X.Y. Li, "Ranking of Closeness Centrality for Large-Scale Social Networks," in F.P. Preparata, X. Wu, J. Yin (eds), *Frontiers in Algorithmics*, FAW 2008, *Lecture Notes in Computer Science*, vol. 5059, Springer, Berlin, Heidelberg, 2008.

[16] J. Golbeck, "Network Structure and Measures," in *Analyzing the Social Web*, Morgan Kaufmann, 2013, pp. 25-44, ISBN 9780124055315.

[17] M. Marchiori, V. Latora, "Harmony in the small-world," *Physica A: Statistical Mechanics and its Applications*, vol. 285, no. 3-4, pp. 539-546, 2000.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:53:46 UTC from IEEE Xplore. Restrictions apply.

Cosmological tests of the dark energy models in Finsler-Randers Space-time

Z. Nekouee,[1,*] Himanshu Chaudhary,[2,3,4,†] S. K. Narasimhamurthy,[5,‡] S. K. J. Pacif,[3,§] and Manjunath Malligawad[5,¶]

[1] *School of Physics, Damghan University, Damghan, 3671641167, Iran*
[2] *Department of Applied Mathematics, Delhi Technological University, Delhi-110042, India*
[3] *Pacif Institute of Cosmology and Selfology (PICS), Sagara, Sambalpur 768224, Odisha, India*
[4] *Department of Mathematics, Shyamlal College, University of Delhi, Delhi-110032, India*
[5] *Department of PG Studies and Research in Mathematics, Kuvempu University, Jnana Sahyadri, Shankaraghatta-577 451, Shivamogga, Karnataka, India*

arXiv:2407.02801v1 [gr-qc] 3 Jul 2024

The Finsler-Randers space-time offers a novel perspective on cosmic dynamics, departing from the constraints of General Relativity. This paper thoroughly investigates two dark energy models resulting from the parametrization of H within this geometric framework. We have conducted some geometrical and physical analysis of the dark energy models in Finslerian geometry. First, We have derived the field equations governing the universe's evolution within the Finsler-Randers formalism, incorporating the presence of dark energy. Through this, we explore its implications on cosmological phenomena, including cosmic expansion, late-time behavior of the universe, cosmological phase transition, and a few more. Also, we employ observational data such as Cosmic Chronometer, Supernovae, Gamma-Ray Bursts, Quasar, and baryon acoustic oscillations to constrain the parameters associated with dark energy in the Finsler-Randers universe. Comparing theoretical predictions with empirical observations, we assess the model viability and discern any deviations from the standard ΛCDM cosmology. Our findings offer intriguing insights into the nature of dark energy within this alternative gravitational framework, providing a deeper understanding of its role in shaping cosmic evolution. The implications of our results extend to fundamental cosmology, hinting at new avenues for research to unravel the mysteries surrounding dark energy and the geometric structure of the universe within non-standard gravitational theories.

I. INTRODUCTION

Dark energy constitutes about 68 % of the total energy content of the universe, a mysterious force thought to be responsible for the universe's accelerated expansion. There are numerous dark energy (DE) models in the literature. However, the most straightforward and widely accepted candidate for DE is Einstein's cosmological constant Λ [1–3]. The Hubble parameter (H) is an essential quantity in cosmology, representing the universe's expansion rate. The concept of cosmic acceleration was initially supported by the findings from observing distant supernovae of type Ia with high redshifts [4, 5]. The motivation for the concept of acceleration stemmed from the independent observations conducted by the supernova teams directed by Perlmutter and Riess, employing diverse methodologies. These observations, along with data gathered from cosmic microwave background (CMB) [6, 7] and studies on large-scale structures [8, 9], contributed to the development of the concept. Researchers provide strong evidence, indicating a more precise comprehension of the universe's accelerating expansion [10–14]. The finding of cosmic acceleration has significant implications for physics and cosmology and has sparked investigations into the basic ideas that underlie cosmological theories. In recent years, many cosmological models have been proposed by researchers, either by changing the gravity theory itself or by modifying the energy-momentum tensor in Einstein's field equations (EFEs) [15–17]. The development of EFEs played a pivotal role in the quest for precise solutions. As a direct outcome, the Schwarzschild exterior solution [18] emerged as the inaugural exact solution. This solution was derived by incorporating the perfect fluid equation of state as a supplementary condition. Efforts in exact solutions development utilizing Einstein's field equations (EFEs) have been concentrated on specific cases, such as solutions for static and spherically symmetric metrics. Numerous noteworthy examples have been derived, including Einstein's static solution, Tolman's solutions [19], the de-Sitter solution [20], Adler's solutions [21], Vaidya and Tikekar solution [22], Knutsen's solutions [23], Buchdahl's solution [24], and Durgapal's solutions [25]. These solutions represent a comprehensive exploration of the possibilities within the framework of EFEs, showcasing the versatility of the equations in capturing various physical scenarios. The observational data is a pivotal inclusion in the cosmology examination, enhancing the credibility and

* zohrehnekouee@gmail.com
† himanshuch1729@gmail.com
‡ nmurthysk@gmail.com
§ shibesh.math@gmail.com
¶ manjunathmalligawad91@gmail.com

applicability of models. Another alternative avenue contributing substantially to modern cosmological investigations is using numerical computation.

In recent advancements in dark energy models, the integration of observational data has significantly improved accuracy in understanding the universe's expansion. A notable contribution to this field is S. K. J. Pacif's [26] study, where two DE cosmological models are investigated. The study involved obtaining exact solutions of the EFEs through cosmological parametrization, specifically utilizing the Hubble parameter (H) within a flat FLRW background. J. K. Singh [27] has investigated a cosmological model within the FLRW framework by employing the Energy Density Scalar Field Differential Equation (EDSFD) parametrization. Constraints derived from observational data are applied to a specific f(Q, T) gravity model [28] characterized by the Hubble parameter transition. Numerical analysis is crucial in investigating the cubic parametrization of the deceleration parameter within the $f(R,T)$ gravitational theory [29]. This study leverages observational data, specifically from $H(z)$ and SNIa datasets, to determine optimal model parameters and the Hubble constant through best-fit values. The brief analysis of three dark energy models, derived from deceleration parameter parametrization [30], reveals noteworthy traits like late-time acceleration and a cosmological transition from early deceleration to late acceleration. Abdulla A. M. [31] investigates the Barrow holographic dark energy model with the Granda–Oliveros infrared cutoff in the recent examination. Ashley Chraya, [32] uses the most recent available observational data to explore the cosmological constraints on the Variable Chaplygin gas model. Vinod Kumar Bhardwaj [33] conducted an investigation into an isotropic and homogeneous cosmological model of the universe within the framework of f(R, T) gravity. Recent studies have delved into various types of dark energy models, including those featuring nonmonotonic energy densities known as omnipotent dark energy (DE) models[34]. Additionally, investigations have been conducted on the Horava-Lifshitz gravity model [35].

In modern theoretical physics, the main goal is to explain how the universe behaves and understand its mysterious components differently. In the last 30 years, observations have indicated that the universe is extending at an accelerated rate, and standard FRW cosmology with ordinary matter fields (cold dark matter, radiation) is not to describe this. The introduction of a cosmological constant Λ in the classical Hilbert action of general relativity (GR) can explain this accelerated expansion. However, its value cannot be calculated theoretically and must be adjusted by hand to match observations. Attempts have been made to relate the cosmological constant to the vacuum energy of quantum fields, but the calculated value deviates from the value suggested by observations. The causality for the accelerated expansion is called "dark energy," although we know very little about it. Several modified gravity theories have been offered as nominees to explain theoretically the observations, usually by modifying the Hilbert action in a Riemannian framework. Finsler Geometry is a hopeful approach to unraveling the secrets of gravity more comprehensively. Instead of using pseudo-Riemannian geometry to describe the gravitational interaction, Finsler space-time geometry offers a viable method for a geometric explanation of the dark matter and dark energy phenomena [36–40]. Within the context of Finsler geometry, various mathematical expressions can be studied, e.g., cosmological implications of scalar-tensor theories were studied in Ref. [41]. These theories were obtained, in effect, from the Lorentz fiber bundle of a Finsler-like geometry. In recent studies, some researchers have investigated dark energy models and used the Finsler geometry as their framework of choice. Finsler and Finsler-like theories take a different approach because they change the geometry instead of the action. Generally, metrical extensions of Riemann geometry can be equipped with a Finsler geometrical structure in a manifold that leads to generalized gravitational field theories. They are drawing inspiration from the osculating Barthel–Kropina geometry [42, 43], Barthel–Randers geometry [44], Finsler–Randers geometry [45, 46] in their investigation. During these years, applications of Finsler geometry, especially the Finsler-Randers (F-R) model, have developed rapidly, mainly in general relativity, astrophysics, and cosmology. The F-R field equations supply a Hubble parameter that contains an extra geometrical term that can be utilized as a probable nominee for dark energy [47].

On the other hand, recent observations of the cosmic microwave background (CMB) by experimentations like the Planck satellite have demonstrated potential anomalies and deviations from statistical isotropy on large angular scales [48, 49]. Noteworthy studies have been executed on the observable anisotropies of the universe [50–53]. These are associated with the very early state of the universe and related to the estimates of the Wiener-filtered map of the cosmic microwave background (CMB), the anisotropic pressure, or the composition of a primordial vector field (instance magnetic field) to the metrical spatial structure of the universe [54–56]. In this case, the form of the scale factor can be impacted by the preliminary field. Lately, anisotropies also have been discussed by the Royal Society, and there is a compelling case that there are anomalies in data [57]. These anomalies, including the well-known low multipole alignments, cold spot, hemispherical asymmetry, and parity asymmetry, have sparked expansive discussions within the cosmology community about their origins and implications. While these anomalies could be mere statistical flukes within the standard Λ-CDM model, they have also been interpreted as potential signatures of new physics or extensions to the standard cosmological framework. One fascinating possibility is that these anomalies might

be related to the primordial anisotropies' existence or preferred directions in the early universe. Such anisotropies could derive from various sources, such as anisotropic inflationary models, vector or tensor fields, or even more exotic scenarios like cosmic birefringence or violations of fundamental symmetries. Probing the effects of such anisotropies on the evolution of the universe and the formation of large-scale structures is an active area of research with potential importance for our understanding of the early universe and the nature of dark energy.

A geometry that possibly connects the Riemannian metric structure of space-time to physical vector fields is a class of the Finsler-Randers variousness of spaces. In these spaces, an electromagnetic field, a magnetic field, or a gauge vector field may derive from the universe's physical source and can be contained in the geometry and create an anisotropic structure [58–62]. As a generalization of Riemannian geometry, Finsler geometry [63, 64] is the most general geometry where the line element depends on space-time coordinates (x) and tangent vectors (y). This y-dependence essentially describes the Finsler field and is merged with the concept of anisotropy, which creates deviations from Riemannian geometry [65]. Therefore, internal variables generate these geometrical anisotropies. With these conditions, Finsler geometry can be regarded as a physical geometry on which the matter dynamics occur, while Riemann geometry is the gravitational geometry [66–68].

II. GRAVITATIONAL FIELD EQUATIONS FOR FINSLER-RANDERS METRIC

The F-R cosmic scenario is established on the Finalerian geometry, which expands the Riemannian geometry. It is worth noticing that Riemannian geometry is also the particular case of the Finslerian geometry. Here, we examine only the main characteristics of the theory (see [63, 68, 69]). Generally, a Finsler space is derived from a generating function F(x, y) on the tangent bundle TM of a manifold M. The generating function F is differentiable on $\widehat{TM} = TM \setminus \{0\}$. The function F is also positively homogeneous of degree one in y. The $n \times n$ Hessian matrix is positive definite at every point of $\widehat{TM}$ that as the following form,

$$g_{ij} = \frac{\partial^2(\frac{1}{2}F^2)}{\partial y^\mu \partial y^\nu} = \frac{1}{2}\dot{\partial}_\mu \dot{\partial}_\nu F^2, \tag{1}$$

In other words, pair (M, F) is called Finsler space-time. In the F-R space-time, we have

$$F = \alpha(x, y) + \beta(x, y), \tag{2}$$

where

$$\alpha(x, y) = \sqrt{a_{ij} y^i y^j}, \qquad \beta(x, y) = \mathfrak{u}_i y^i, \tag{3}$$

$a_{ij}(x)$ is the FRW metric defined as a tensor corresponding to the Riemannian metric,

$$a_{ij}(x) = diag\left(1, -\frac{a^2(t)}{1-\kappa r^2}, -a^2(t)r^2, -a^2(t)r^2 \sin^2\theta\right), \tag{4}$$

and $\kappa = 0, \pm 1$ for a flat, closed and hyperbolic geometry respectively. Consequently, the geometrical structure of Finsler-Randers space is a generalization of Riemannian geometry. Here, the spatial coordinates are comoving the time coordinate defines the proper time investigated by the comoving observer. The vector $y^i = \frac{dx^i}{d\tau}$ expresses the four tangent vectors that depict the velocity of a comoving observer towards an appropriate family of world lines, namely fluid flow pathways, in a locally anisotropic universe. The arclength τ is the proper time. So we have $y^i = (1, 0, 0, 0)$. $\mathfrak{u}_i$ expresses a weak primordial vector field with $|\mathfrak{u}_i| \ll 1$ and $\mathfrak{u}_i = (\mathfrak{u}_0, 0, 0, 0)$, where $\mathfrak{u}_0 = \mathfrak{u}_0(t)$. Here, β is an 1-form. Now, using Eq. (1), the metric for the Finslerian space-time can be defined as follows [47, 70]:

$$g_{ij} = \mathfrak{g}_{ij} + \frac{1}{4\alpha}\left(\mathfrak{u}_i y_j + \mathfrak{u}_j y_i\right) - \frac{\beta}{\alpha^3} y_i y_j + \mathfrak{u}_i \mathfrak{u}_j, \tag{5}$$

where

$$\mathfrak{g}_{ij} = \frac{F}{\alpha}(x, y) a_{ij}(x). \tag{6}$$

Cartan tensor can be easily defined in full expression as follows [71]

$$C_{ijk}=\frac{1}{2}\left[\frac{1}{\alpha}\mathfrak{S}_{(ijk)}(a_{ij}\mathfrak{u}_k)-\frac{1}{\alpha^3}\mathfrak{S}_{(ijk)}(y_iy_j\mathfrak{u}_k)-\frac{\beta}{\alpha^3}\mathfrak{S}_{(ijk)}(a_{ij}y_k)+\frac{3\beta}{\alpha^5}y_iy_jy_k\right], \tag{7}$$

where $\mathfrak{S}_{(ijk)}$ expresses the sum over the cyclic permutation of the indices. It has been found that $C_{000}=\frac{\mathfrak{u}_0}{2}$ [47]. The Finsler–Randers field equations are the following form,

$$L_{ij}=\frac{8\pi G}{c^4}\left(T_{ij}-\frac{1}{2}T\mathfrak{g}_{ij}\right), \tag{8}$$

where

- T_{ij} are the energy-momentum tensor components for describing Finslerian anisotropic matter

$$T_{ij}(x,y(x))=(\rho+P)y_i(x)y_j(x)-P\mathfrak{g}_{ij}(x,y(x)), \tag{9}$$

  and T is the trace of them. Thus we have

$$T_{00}=\rho,\quad T_{11}=-P\mathfrak{g}_{11},\quad T_{22}=-P\mathfrak{g}_{22},\quad T_{33}=-P\mathfrak{g}_{33},\quad T=\rho\frac{\alpha}{F}-3P, \tag{10}$$

  where $\frac{F}{\alpha}\approx 1$ at the weak field limit.

- $L_{ij}=L^k_{ikj}$ is Ricci tensor and by inflicting the conditions $\ddot{\mathfrak{u}}_0,\dot{\mathfrak{u}}_0^2\approx 0$, its non-zero components as the following form [72],

$$L_{00}=-3\left(\frac{\ddot{a}}{a}+\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0\right), \tag{11}$$

$$L_{11}=\frac{a\ddot{a}+2\dot{a}^2+2\kappa+4a\dot{a}\dot{\mathfrak{u}}_0}{1-\kappa r^2}, \tag{12}$$

$$L_{22}=\left(a\ddot{a}+2\dot{a}^2+2\kappa+4a\dot{a}\dot{\mathfrak{u}}_0\right)r^2, \tag{13}$$

$$L_{33}=\left(a\ddot{a}+2\dot{a}^2+2\kappa+4a\dot{a}\dot{\mathfrak{u}}_0\right)r^2\sin^2\theta. \tag{14}$$

Now by imposing $c=1$ and $\kappa=0$ we have

$$\frac{\ddot{a}}{a}+2\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0=\frac{-4\pi G}{3}(\rho+3P), \tag{15}$$

$$\frac{\ddot{a}}{a}+2\frac{\dot{a}^2}{a^2}+4\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0=4\pi G(\rho-P), \tag{16}$$

where the symbol dot denotes the derivative with respect to t and $\dot{\mathfrak{u}}_0<0$ [73]. By using Eqs. (15) and (16) we obtain the Friedmann equation

$$\frac{\dot{a}^2}{a^2}+\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0=\frac{8\pi G}{3}\rho, \tag{17}$$

and

$$-2\frac{\ddot{a}}{a}-\frac{\dot{a}^2}{a^2}-5\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0=8\pi GP. \tag{18}$$

Eq. (17) is similar to the Friedmann equation in the Riemannian framework, apart from the extra term $\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0$. We associate this extra term with the present universe's anisotropy. By using Eqs. (17) and (18) we obtain deformed parameters as following form

$$\tilde{H}^2=\frac{\dot{a}^2}{a^2}+\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0,\quad\xrightarrow{\dot{\mathfrak{u}}_0^2\approx 0}\quad\tilde{H}=\left(\frac{\dot{a}}{a}+\frac{\dot{\mathfrak{u}}_0}{2}\right),$$

$$\tilde{q}=-\left(\frac{\ddot{a}}{a}+\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0\right)\tilde{H}^{-2},$$

Parameter	Classification
$\omega_{\mathfrak{de}} = -1$	cosmological constant
$\omega_{\mathfrak{de}} =$ Constant$\neq -1$	cosmic strings, domain walls, etc.
$\omega_{\mathfrak{de}} \neq$Constant	scalar fields (quintessence, k-essence etc.), braneworlds, Dirac-Born-Infeld (DBI) action, Chaplygin gas etc.
$\omega_{\mathfrak{de}} < -1$	phantom models

TABLE I: Classification of ω_{de} parameter.

where $\tilde{H}$ and $\tilde{q}$ are Hubble parameter and deceleration parameter, respectively. Conservation of energy-momentum results from Eqs. (17) and (18) as follows,

$$8\pi G\dot{\rho} + 24\pi G(\rho + P)H + 6H^2\dot{\mathfrak{u}}_0 + 12\pi G(\rho + P)\dot{\mathfrak{u}}_0 = 0. \tag{19}$$

The continuity Eq (19) plays an essential role in evolution because it deals with matter and its interactions. In current cosmology, two types of dark energy models are generally debated. Dark energy interaction models (taking into account the interaction between cold dark matter and dark energy) [74] and non-interaction models of dark energy, where matters are authorized to evolve individually [75]. Up to date, there are no known interactions other than gravity between the matter and dark energy. The present study only refers to non-interacting models. The system of equations is non-linear ordinary differential equations, and finding exact solutions is difficult. In the past, many attempts have been made to find exact and numerical solutions for EFEs. In the next section, solving techniques from the above system of equations will be discussed in detail.

III. EINSTEIN FIELD EQUATIONS FOR FINSLER-RANDERS METRIC WITH DARK ENERGY

The nature of dark energy and its candidature are not yet known, and its expression as a source term in Einstein's field equations is speculative. However, DE is assumed to be homogeneous, permeating all over the space, and the energy-momentum tensor can be expressed as a perfect fluid $T^{\mathfrak{de}}_{ij}$. In this case, its equation of state is in the form $P_{\mathfrak{de}} = \omega_{\mathfrak{de}}\rho_{\mathfrak{de}}$ where $\omega_{\mathfrak{de}}$ is the equation of state (EoS) parameter and is a function of time. Generally, it holds for the inequality $\omega_{\mathfrak{de}} < 0$. The appropriate value of $\omega_{\mathfrak{de}}$ is hotly debated, and analysis of some observational data suggests that its value is -1.018 ± 0.031 (Planck 2018 or SNe IA^a) [76]. Different values of $\omega_{\mathfrak{de}}$ in particular ranges produce different candidates, and they can be generally classified as seen in Table I, See [77, 78] (and Refs. therein) for a more extensive list of dark energy models. In each of them, fascinating cases exist that are associated with problems. For example, the cosmological constant is the most consistent model for dark energy that explains the observations but faces fine-tuning issues. Similarly, phantom models are fascinating, where the weak energy condition ($\rho > 0$, $P + \rho > 0$) with finite time singularity property is violated. In general relativity, dark energy can be introduced by adding the energy-momentum tensor $T^{\mathfrak{de}}$ as a perfect fluid together with the matter source $T^{\mathfrak{m}}$ into the Einstein field equations $G_{ij} = 8\pi G T_{ij}$ as following form,

$$T^{\mathfrak{tot}} = T^{\mathfrak{de}} + T^{\mathfrak{m}}, \tag{20}$$

$\rho_{\mathfrak{tot}} = \sum \rho + \rho_{\mathfrak{de}}$ and $P_{\mathfrak{tot}} = \sum P + P_{\mathfrak{de}}$ denote total energy densities and the total pressure corresponding to all kinds of matter (baryonic matter, dark matter, and radiation) and dark energy, respectively. According to Eq. (20), we can rewrite Eqs. (17), (18), and (19) as follows,

$$\frac{\dot{a}^2}{a^2} + \frac{\dot{a}}{a}\dot{\mathfrak{u}}_0 = \frac{8\pi G}{3}\rho_{\mathfrak{tot}}, \tag{21}$$

$$-2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - 5\frac{\dot{a}}{a}\dot{\mathfrak{u}}_0 = 8\pi G P_{\mathfrak{tot}}, \tag{22}$$

$$8\pi G\dot{\rho}_{\mathfrak{tot}} + 24\pi G(\rho_{\mathfrak{tot}} + P_{\mathfrak{tot}})H + 6H^2\dot{\mathfrak{u}}_0 + 12\pi G(\rho_{\mathfrak{tot}} + P_{\mathfrak{tot}})\dot{\mathfrak{u}}_0 = 0. \tag{23}$$

The system of the above equations has only two independent equations with five unknowns a, ρ, P, $\rho_{\mathfrak{de}}$, $P_{\mathfrak{de}}$ (or $\omega_{\mathfrak{de}}$). The universe's large-scale homogeneous distribution of matter requires the barotropic equation of state to be considered $P = \omega\rho$, $\omega \in [0, 1]$. The equation of state expresses various kinds of matter sources in the universe depending on the discrete or dynamical values of the equation of state parameter as follows: baryonic matter ($\omega_{\mathfrak{bm}} = 0$), dark matter ($\omega_{\mathfrak{dm}} = 0$), radiation ($\omega_{\mathfrak{r}} = \frac{1}{3}$), stiff matter ($\omega_{\mathfrak{s}} = 1$), quintessence ($-1 < \omega_{\mathfrak{q}} < -\frac{1}{3}$), etc.

This extra equation represents the third restriction equation. Another restriction equation can be the regard of the EoS of dark energy ($\omega_{\mathfrak{de}}$ = constant or a function of time t or function of the scale factor or redshift), which is well known as a parametrization of dark energy EoS. These four equations can describe the cosmological dynamics of the universe where all the geometrical or physical parameters such as Hubble parameter (H), deceleration parameter (q), jerk parameter (j), densities (ρ , $\rho_{\mathfrak{de}}$), pressures (P, $P_{\mathfrak{de}}$), EoS parameter ($\omega_{\mathfrak{de}}$), density parameter (Ω_i), etc. are signified as functions of either scale factor or the redshift $z = \frac{a_0}{a} - 1$ (a_0 is the current value of the scale factor, usually normalized to $a_0 = 1$). But, we still need another equation to solve the system. The time evolution of the scale factor a is not yet known. Various schemes exist for parametrizing the scale factor and its higher-order derivatives (H, q, j, etc.). They provide the complete solution of EFEs, i.e., the explicit forms of the cosmic parameters as a function of t. An essential analysis of the solution methods of EFEs in general relativity theory or modified theories has two parts: One is the parametrization of geometrical parameters, and another is the parametrization of the physical parameters. The first type of parameterization scheme (geometric parameterization) is considered to find exact solutions that discuss the expansion dynamics of the universe and provide the time evolution of the physical parameters. This method is generally known as an independent model for cosmological models or cosmological parameterization [79–81]. It does not affect the background theory and provides solutions for EFEs. As well as it has the advantage of reconstructing the cosmological history of the universe and describing some of the phenomena of the universe. The goal is to achieve the exact solution of the Einstein field equations in standard general relativity with a simple parametrization of the Hubble parameter under the Finsler-Randers structure.

IV. PARAMETRIZATION OF HUBBLE PARAMETER IN FINSLER-RANDERS MODEL

Cosmological investigations provide clues to explore the evolution of the observable universe in a model-independent way in terms of kinematic variables [82]. Furthermore, the cosmological parameters analysis helps to study dark energy without any assumptions of any particular cosmological model other than the cosmological principle. In this paper, we intend to use a simple parametrization of the Hubble parameter, which is an explicit function of the cosmic time "t" as following form [81],

$$H(t) = \frac{k_2 t^{\mathfrak{m}}}{(t^{\mathfrak{n}} + k_1)^{\mathfrak{p}}}, \tag{24}$$

where $\mathfrak{m}, \mathfrak{n}, \mathfrak{p} \in \mathbb{R}$, k_1, $k_2 \neq 0$, and both have the dimensions of time. Some special values of the parameters $\mathfrak{m}$, $\mathfrak{n}$, and $\mathfrak{p}$ present some distinct models [81]. The parameterization (24) generalizes several well-known models, e.g., ΛCDM model, power law model, hybrid expansion model, jump model, linear variable deceleration parameter model, etc. As we know, $|\mathfrak{u}_i| \ll 1$ for parametrization Hubble parameter in Finsler structure we take $\mathfrak{u}_0 = \exp(-t)$. Table II can describe cosmic phase transition phenomena for negative values of k_1, and k_2 in two models. In studying late-time

Models	$H(t)$	$\tilde{H}(t)$	$a(t)$	$\tilde{a}(t)$
Model 1	$\frac{k_2}{t(k_1-t)}$	$\frac{k_2}{t(k_1-t)} - \frac{1}{2}e^{-t}$	$\beta\left(\frac{t}{k_1-t}\right)^{\frac{k_2}{k_1}}$	$\beta\left(\frac{t}{k_1-t}\right)^{\frac{k_2}{k_1}} e^{\frac{1}{2}e^{-t}}$
Model 2	$\frac{k_2}{t(k_1-t^2)}$	$\frac{k_2}{t(k_1-t^2)} - \frac{1}{2}e^{-t}$	$\beta\left(\frac{t^2}{k_1-t^2}\right)^{\frac{k_2}{2k_1}}$	$\beta\left(\frac{t^2}{k_1-t^2}\right)^{\frac{k_2}{2k_1}} e^{\frac{1}{2}e^{-t}}$

TABLE II: The Models

universe and observational studies, expressing all the cosmological parameters as functions of redshift z is convenient. It is well known that the Hubble parameter ($H = \frac{\dot{a}}{a}$) directly probes the universe's expansion history where $\dot{a}$ is the rate of change of the universe's scale factor. The Hubble parameter is also related to the differential redshift $H(z) = \frac{1}{z+1}\frac{dz}{dt}$, where dz is obtained from spectroscopic surveys. Table III shows the t-z relations, that $\alpha = \frac{k_1}{k_2}$ and parameters α and β are sufficient to describe these models.

Models	$t(z)$	$\tilde{H}(z)$
Model 1	$\frac{k_1}{1+[\beta(1+z)]^{\alpha}}$	$\frac{H_0(1+[\beta(1+z)]^{\alpha})^2}{(1+\beta^{\alpha})^2(1+z)^{\alpha}} - \frac{1}{2}e^{-\frac{(1+\beta^{\alpha})^2}{\alpha H_0 \beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}$
Model 2	$\frac{\sqrt{k_1}}{\sqrt{1+[\beta(1+z)]^{2\alpha}}}$	$\frac{H_0\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{3}{2}}}{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}} - \frac{1}{2}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0 \beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}$

TABLE III: Hubble parameter vs. z in two models.

V. METHODOLOGY

In cosmology, parameter estimation often employs the Bayesian framework, which entails calculating the posterior distribution of parameters θ given the observed data D. This relationship is represented as:

$$P(\theta \mid D) = \frac{\mathcal{L}(D \mid \theta)P(\theta)}{P(D)}, \tag{25}$$

where $\mathcal{L}(D \mid \theta)$ is the likelihood function, $P(\theta)$ denotes the prior distribution of the parameters, and $P(D)$ is the marginal likelihood. The posterior distribution is explored across the parameter space θ, typically using algorithms such as Metropolis-Hastings. This algorithm directs a random walker through the parameter space, favoring areas with higher likelihood values. Parameter estimation involves determining the mean and uncertainty of the parameter distribution by examining the locations where the walker frequently visits and the extent of its deviations within the parameter space. To approximate the shape of the posterior distribution, the walker needs to take a sufficiently large number of steps. In Bayesian model selection, the evidence for each model is computed individually and then compared. This task usually demands substantial computational resources and often utilizes nested sampling methods. However, simpler tools like information criteria can offer reliable insights into model preferences when the posterior distributions closely resemble a Gaussian distribution. Our study investigates Dark Energy Models within Finsler geometry and Riemannian geometry frameworks. We aim to determine the optimal parameter values for these models and estimate the present-day Hubble constant (H_0). We employ Markov Chain Monte Carlo (MCMC) analysis using the Polychord package to explore the parameter space and obtain reliable estimates efficiently. This approach ensures robust exploration of the parameter space and accurate parameter estimation.

A. Cosmological Data Sets

1. *Cosmic Chronometers (CC)*

Cosmic chronometers are reliable probes for studying cosmic expansion, providing a model-independent method to measure the Hubble constant and the expansion rate. By analyzing the age and metallicity of nearby passive galaxies, we can estimate the expansion rate $H_{\rm CC}(z)$ at a given redshift $z_{\rm CC}$. This estimation is based on the observation that the expansion rate can be approximated by the ratio of the redshift difference to the time difference, adjusted by the redshift itself: $H_{\rm CC}(z) \approx -(\Delta z_{\rm CC}/\Delta t)/(1+z_{\rm CC})$. We have gathered cosmic chronometer data from various sources [83–86] covering a redshift range of approximately $0.07 \lesssim z \lesssim 1.97$. These data provide direct constraints on the universe's expansion history. To evaluate the consistency between theoretical predictions and cosmic chronometer measurements across different redshifts, we calculate the chi-squared distance $\chi^2_{\rm CC}$:

$$\chi^2_{\rm CC} = \sum_{i=1}^{31} \left(\frac{H(z_{\rm CC}) - H_{\rm CC}(z_{\rm CC})}{\sigma_{\rm CC}(z_{\rm CC})} \right)^2. \tag{26}$$

In this equation, $H_{\rm CC}(z_{\rm CC}) \pm \sigma_{\rm CC}(z_{\rm CC})$ represents the cosmic chronometer measurements of the expansion rate at redshift $z_{\rm CC}$, along with their uncertainties. We assume that the cosmic chronometer observations are uncorrelated.

2. *Type Ia Supernova (SNIa)*

One of the most renowned and commonly utilized cosmological tools is the study of distant Type Ia Supernovae. These supernova explosions are exceptionally luminous, often rivaling the brightness of their host galaxies [87]. Their

observed light curves typically exhibit peak brightness levels that remain relatively unaffected by distance, making them invaluable as standard candles. A key metric used to constrain cosmological models is the observed distance modulus, denoted as $\mu_{\rm obs}$. For our analysis, we use the most recent dataset available, the binned Pantheon dataset described in [87]. The likelihood function for the Type Ia Supernovae data, $L_{\rm SNIa}(Y;M)$, is expressed as:

$$L_{\rm SNIa}(Y;M) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{40} m_i C_{\rm cov}^{-1}\,{\rm cov}_m^{\dagger}\, i\right). \tag{27}$$

In this context, Y represents the vector of free parameters of the cosmological model. The term m_i is defined as the difference between the observed distance modulus $\mu_{{\rm obs},i}$ and the theoretical distance modulus $\mu_{\rm theor}(z_i)$, adjusted by M. The theoretical distance modulus $\mu_{\rm theor}$ is calculated using the standard luminosity distance formula:

$$\mu_{\rm theor} = 5\log\left(\frac{D_L}{1\,{\rm Mpc}}\right) + 25, \tag{28}$$

where D_L represents the luminosity distance, given by $D_L = c(1+z)\int_0^z \frac{dz'}{H(z')}$. This formula is applicable in a flat Friedmann-Robertson-Walker (FRW) space-time, irrespective of the underlying cosmology. Finally, $C_{\rm cov}$ denotes the covariance matrix associated with the binned Pantheon dataset. The parameter M is an intrinsic free parameter within the Pantheon dataset, capturing various observational uncertainties, including those related to host galaxy properties. To broaden our observational spectrum, we include 24 binned quasar distance modulus data from [88], a sample of 162 Gamma-Ray Bursts (GRBs) as described in [89]

3. *Baryon Acoustic Oscillations (BAO)*

For our analysis, we utilize a comprehensive list of BAO data points, including effective redshifts ($z_{\rm eff}$), observables, measurements, and associated errors, as provided in [90]. The BAO feature in the transverse direction offers a measurement of $\frac{D_H(z)}{r_d} = \frac{c}{H(z)r_d}$, where $D_H(z)$ is the Hubble distance at redshift z, and r_d is the sound horizon at the drag epoch. In a flat Universe, the comoving angular diameter distance is given by $D_M = \frac{c}{H_0}\int_0^z \frac{dz'}{E(z')},$. where $E(z)$ is the dimensionless Hubble parameter. The angular diameter distance D_A is defined as $D_A = \frac{D_M}{1+z}$, and the volume-averaged distance $D_V(z)/r_d$ contains information about the BAO peak coordinates. The volume-averaged distance is given by$D_V(z) = \left[zD_H(z)D_M^2(z)\right]^{1/3},$. with r_d being the sound horizon at the drag epoch, measured as 147.1 Mpc [91]. To compare theoretical predictions with observational BAO data, we calculate the distance function $\chi^2_{\rm BAO}$ as follows:

$$\chi^2_{\rm BAO} = \sum_{i=1}^{17}\left(\frac{H(z_{\rm BAO}) - H_{\rm BAO}(z_{\rm BAO})}{\sigma_{\rm BAO}(z_{\rm BAO})}\right)^2, \tag{29}$$

where $H_{\rm BAO}(z_{\rm BAO}) \pm \sigma_{\rm BAO}(z_{\rm BAO})$ represents the BAO-derived measurements of the expansion rate at redshift $z_{\rm BAO}$, scaled by the sound horizon radius. In our analysis, we have employed a nested sampler, implemented within the open-source package Polychord [92], in conjunction with the GetDist package [93] for presenting the results. Figs 1 and 2 depict the 68% and 95% confidence levels for key cosmological parameters in both the Riemannian and Finslerian dark energy models. Table IV presents the best-fit values of the model parameters α and β, along with the current Hubble constant H_0.

VI. OBSERVATIONAL AND THEORETICAL COMPARISONS OF THE HUBBLE PARAMETER, HUBBLE DIFFERENCE, AND APPARENT MAGNITUDE

After obtaining the best-fit values for both dark energy parametrizations in Riemannian and Finslerian models, it is necessary to compare these models with real observational datasets and the standard ΛCDM paradigm. To conduct a comprehensive comparative study, we will plot the Hubble function for all three scenarios: the Riemannian model, the Finslerian model, and the ΛCDM model. We will use the 31 Cosmic Chronometer (CC) datasets using their corresponding error bars for this analysis. Additionally, we will plot the difference between the Hubble functions of the Riemannian and Finslerian models relative to the ΛCDM model, denoted as $H_{\rm Model}(z) - H_{\Lambda{\rm CDM}}(z)$, against the CC datasets. Furthermore, we will plot the apparent magnitude in each mode and compare these with each other

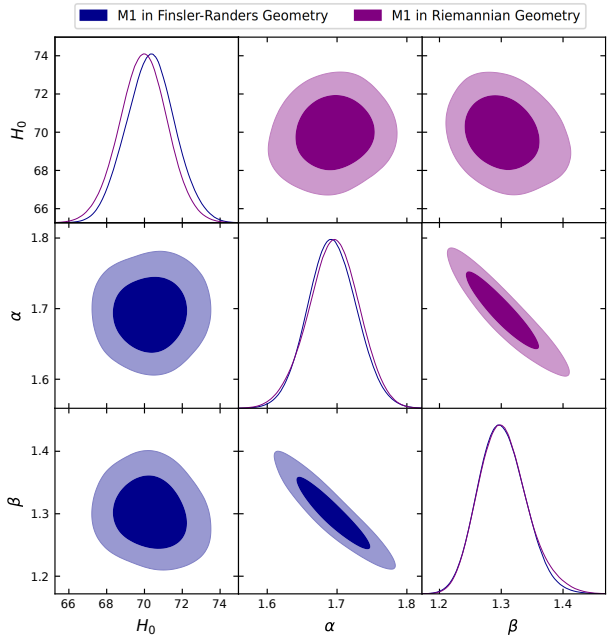


FIG. 1: MCMC confidence contours at 1σ and 2σ for Model 1.

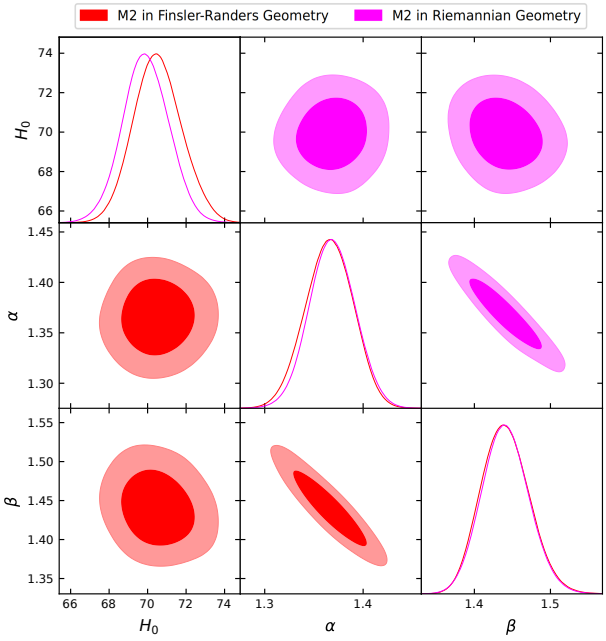


FIG. 2: MCMC confidence contours at 1σ and 2σ for Model 2.

MCMC Results					
Finslerian Geometry	Parameters	Best-fit Value	Riemannian Geometry	Parameters	Best-fit Value
Model 1	H_0	$70.326621^{\pm 2.380077}_{\pm 1.381923}$	Model 1	H_0	$69.946236^{\pm 2.522770}_{\pm 1.203929}$
	α	$1.693215^{\pm 0.066342}_{\pm 0.033153}$		α	$1.695306^{\pm 0.073840}_{\pm 0.033177}$
	β	$1.300786^{\pm 0.068124}_{\pm 0.037316}$		β	$1.302990^{\pm 0.069706}_{\pm 0.040574}$
Model 2	H_0	$70.517486^{\pm 2.347270}_{\pm 1.219293}$	Model 2	H_0	$69.901832^{\pm 2.333254}_{\pm 1.073915}$
	α	$1.365617^{\pm 0.046564}_{\pm 0.024308}$		α	$1.368668^{\pm 0.043644}_{\pm 0.022250}$
	β	$1.440327^{\pm 0.055222}_{\pm 0.031578}$		β	$1.440785^{\pm 0.058645}_{\pm 0.029173}$

TABLE IV: Summary of MCMC Results.

and with actual observational data. This comparison is essential for assessing the viability and accuracy of both dark energy parametrizations in explaining the Universe's accelerated expansion. Analyzing the Hubble function allows us to directly examine the rate of expansion over time while comparing $H_{\rm Model}(z) - H_{\Lambda{\rm CDM}}(z)$ helps us understand deviations from the standard model. The apparent magnitude comparison provides insights into the models' predictions of luminosity distance, which is crucial for understanding cosmological distances and the scale of the Universe.

A. Hubble Parameter $H(z)$

Figs. 3 and 4 illustrate the evolution of the Hubble parameters as a function of redshift for both the Riemannian and Finslerian models, with the ΛCDM model (black line) included for comparison, using cosmological parameters $\Omega_{\rm m0} = 0.3$ and $\Omega_\Lambda = 0.7$. In Fig. 3 for model M1, a deviation between the Riemannian (purple line) and Finslerian

(blue line) models and the ΛCDM model is evident at $z > 0.6$. However, all three models exhibit close agreement at $z < 0.6$. Similarly, in Fig. 4 for model M2, a deviation between the Riemannian (magenta line) and Finslerian (red line) models and the ΛCDM model is observed at $z > 1.2$ and close agreement at $z < 1.2$.

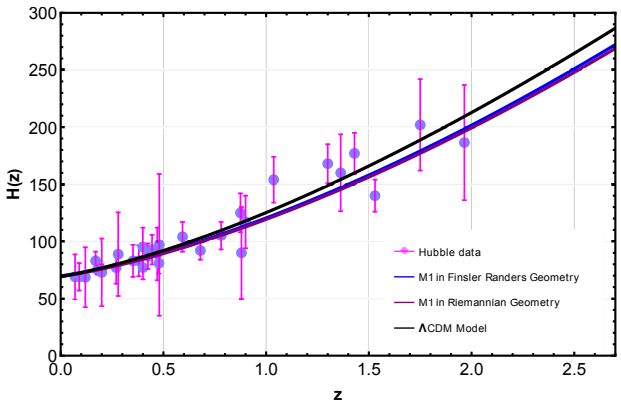


FIG. 3: Evolution of the Hubble Parameter as a function of redshift in Riemannian and Finslerian Modes

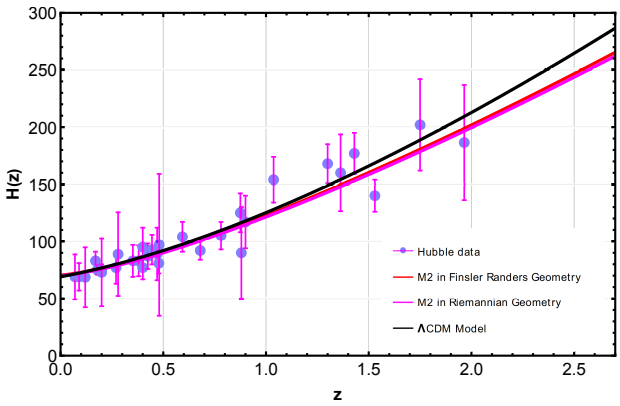


FIG. 4: Evolution of the Hubble Parameter as a function of redshift in Riemannian and Finslerian Modes

B. Hubble Difference $\Delta H(z)$

Figs. 5 and 6 illustrate the evolution of the difference between the Hubble parameter H predicted by specific cosmological models, Riemannian and Finslerian in our case and the Hubble parameter H predicted by the ΛCDM model, both evaluated at a given redshift z. Fig. 5 shows $\Delta H(z) = H_{\mathrm{M1}}(z) - H_{\Lambda\mathrm{CDM}}(z)$. It can be observed that $\Delta H(z)$ is significant at $z > 0.6$ for both the Riemannian (purple line) and Finslerian (blue line) models compared to the ΛCDM model. However, at $z < 0.6$, $\Delta H(z)$ becomes negligible. Fig. 6 for model M2 shows $\Delta H(z)$. A significant difference between the Riemannian (magenta line) and Finslerian (red line) models and the ΛCDM model is observed at $z > 1$. However, $\Delta H(z)$ becomes negligible at $z < 1$.

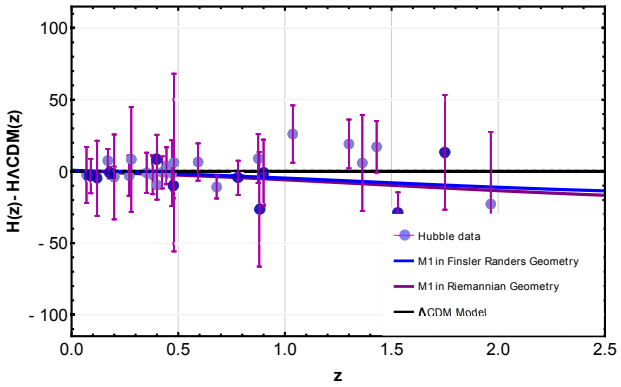


FIG. 5: Evolution of the Hubble Difference in Riemannian and Finslerian Models

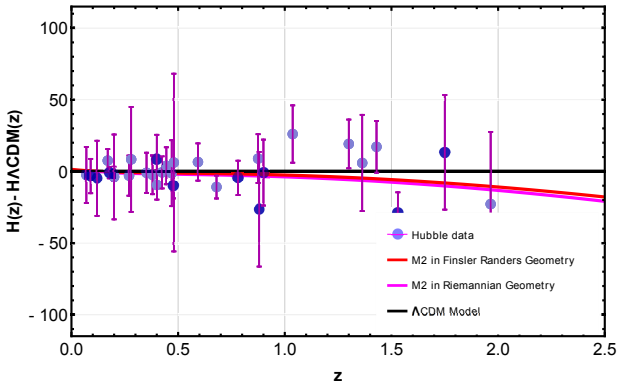


FIG. 6: Evolution of the Hubble Difference in Riemannian and Finslerian Models.

C. Distance Modulus $\mu(z)$

Figs. 7 and 8 depict the evolution of the distance modulus $\mu(z)$ for a Type Ia supernova (SINa), Gamma Ray Burst (GRB), and Quasar (Q). The distance modulus is an astronomical measure used to quantify the apparent brightness of celestial objects based on their redshift z, which indicates how much the Universe has expanded since the object's

light was emitted. It is related to the luminosity distance D_L by the formula: $\mu = 5\log_{10}(D_L) + 25$, where μ is the distance modulus and D_L is the luminosity distance in megaparsecs (Mpc). In the standard cosmological model, the luminosity distance is calculated as: $D_L = (1+z) \cdot D_C(z)$. Here, z is the redshift, and $D_C(z)$ is the comoving distance, which represents the spatial distance between the observer and the object, accounting for the expansion of the Universe. The comoving distance $D_C(z)$ is given by an integral involving the inverse of the Hubble parameter $H(z)$: $D_C(z) = \int_0^z \frac{c\,dz'}{H(z')}$, where c is the speed of light and $H(z)$ is the Hubble parameter at redshift z. Fig. 7 illustrates a close agreement between the Riemannian (purple line) and Finslerian (blue line) modes with ΛCDM model and SNIa dataset at low redshift. However, differences become evident at high redshifts when considering the SNIa dataset. Fig. 8 exhibits a similar trend in model M2, where both Riemannian (magenta line) and Finslerian (red line) modes closely resemble the ΛCDM model.

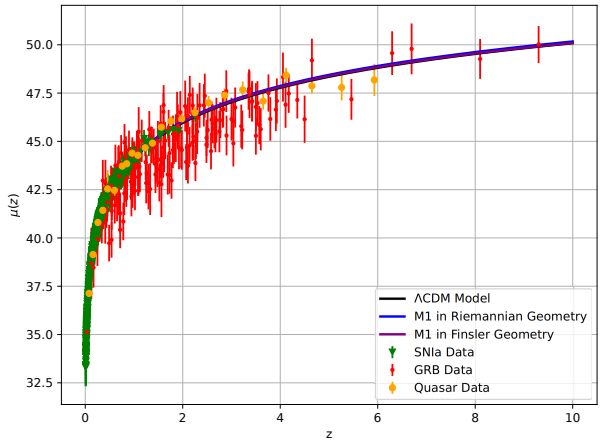


FIG. 7: The evolution of Distance Modulus in Riemannian and Finslerian modes.

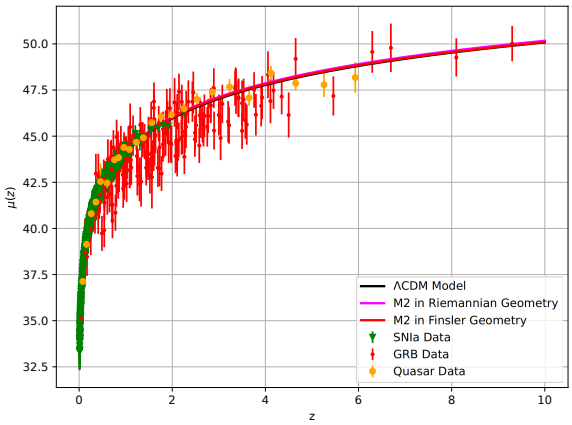


FIG. 8: The evolution of Distance Modulus in Riemannian and Finslerian modes.

VII. COSMOGRAPHIC PARAMETER

A. Deceleration Parameter

The deceleration parameter (DP), q, is a dimensionless measure in cosmology that describes the rate of change of the universe's expansion. It is defined as $q = -\frac{\ddot{a}a}{\dot{a}^2}$, where a is the scale factor of the universe, $\dot{a}$ is its first derivative with respect to time (the expansion rate), and $\ddot{a}$ is its second derivative (the acceleration of the expansion). A positive q indicates a decelerating expansion, while a negative q indicates an accelerating expansion. Understanding q is crucial for studying the dynamics of the universe, including the transition from a decelerated to an accelerated phase, which is associated with the influence of dark energy. In the case of Finslerian Geometry, the expression of the deceleration parameter in terms of redshift z for Models 1 and 2 is provided in Table V. Figs 9 and 10 show the evolution of the deceleration parameter for Models 1 and 2 in both Riemannian and Finslerian geometries, compared with the ΛCDM model.

B. Jerk and Snap Parameters

As we know, $q(t)$ represents the deceleration parameter, which measures the Universe's acceleration. Now, we are going to introduce other parameters such as $j(t) = \frac{\dddot{a}}{aH^3}$ (jerk parameter), and $s(t) = \frac{a^{(4)}}{aH^4}$ (snap parameter). These parameters play significant roles in the cosmographic analysis of the Universe, specifically in understanding the

Models	$t(z)$	$\tilde{q}(z)$
Model 1	$\frac{k_1}{1+[\beta(1+z)]^{\alpha}}$	$\frac{\frac{(1+\beta^{\alpha})^2(1+z)^{\alpha}}{H_0(1+[\beta(1+z)]^{\alpha})^2}e^{-\frac{(1+\beta^{\alpha})^2}{\alpha H_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}-1+\alpha-2\alpha(1+[\beta(1+z)]^{\alpha})^{-1}}{\frac{-(1+\beta^{\alpha})^2(1+z)^{\alpha}}{H_0(1+[\beta(1+z)]^{\alpha})^2}e^{-\frac{(1+\beta^{\alpha})^2}{\alpha H_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}+1}$
Model 2	$\frac{\sqrt{k_1}}{\sqrt{1+[\beta(1+z)]^{2\alpha}}}$	$\frac{\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}}{H_0\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{3}{2}}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}-1+\alpha-\frac{3\alpha}{1+[\beta(1+z)]^{2\alpha}}}{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}}{H_0\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{3}{2}}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}+1}$

TABLE V: Expression of the deceleration parameter

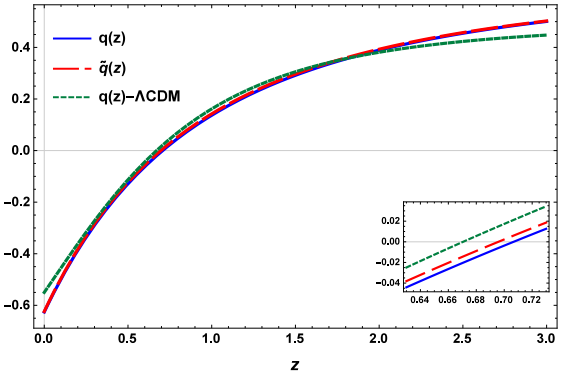


FIG. 9: The evolution of deceleration parameter from past ($z = 4$) to far future with a phase transition in Riemannian and Finslerian modes.

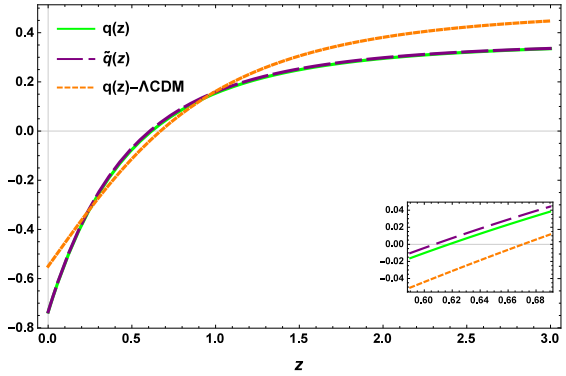


FIG. 10: The evolution of deceleration parameter from past ($z = 4$) to far future with a phase transition in Riemannian and Finslerian modes.

behavior of acceleration, jerk, and snap, and help distinguish between different dark energy models. Table VI displays the mathematical expressions of the jerk parameters in terms of redshift.

- Anisotropic jerk parameter: $\tilde{j}(t) = \frac{\dddot{a}+\frac{3}{2}\ddot{a}\dot{u}_0}{a\tilde{H}^3}$,

- Anisotropic snap parameter: $\tilde{s}(t) = \frac{a^{(4)}+2\,\dddot{a}\dot{u}_0}{a\tilde{H}^4}$,

Models	$\tilde{j}(z)$
Model 1	$\frac{-3\frac{(1+\beta^{\alpha})^2(1+z)^{\alpha}}{H_0(1+[\beta(1+z)]^{\alpha})^2}e^{-\frac{(1+\beta^{\alpha})^2}{\alpha H_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}\left(\frac{\alpha}{1+\beta^{\alpha}(1+z)^{\alpha}}-\frac{\alpha}{2}+\frac{1}{2}\right)+1+\alpha(2\alpha-3)+\frac{6\alpha}{1+\beta^{\alpha}(1+z)^{\alpha}}\left[1-\alpha+\frac{\alpha}{1+\beta^{\alpha}(1+z)^{\alpha}}\right]}{-\frac{3}{2}\frac{(1+\beta^{\alpha})^2(1+z)^{\alpha}}{H_0(1+[\beta(1+z)]^{\alpha})^2}e^{-\frac{(1+\beta^{\alpha})^2}{\alpha H_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}+1}$
Model 2	$\frac{\frac{-9}{2}\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}}{H_0\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{3}{2}}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}\left(\frac{\alpha}{1+[\beta(1+z)]^{2\alpha}}-\frac{\alpha}{3}+\frac{1}{3}\right)+1-3\alpha+2\alpha^2+\frac{12\alpha^2}{\left(1+[\beta(1+z)]^{2\alpha}\right)^2}+\frac{3\alpha(3-2\alpha)}{1+[\beta(1+z)]^{2\alpha}}}{-\frac{3}{2}\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}}{H_0\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{3}{2}}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}+1}$

TABLE VI: Expression of the jerk parameter.

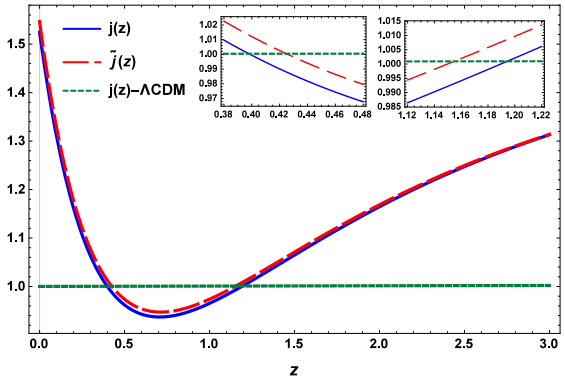


FIG. 11: Jerk parameter vs. z in Riemannian and Finslerian modes.

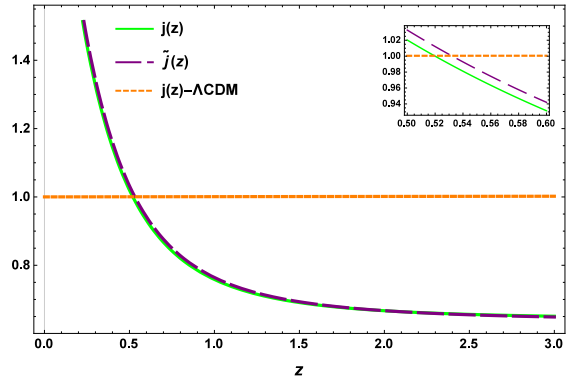


FIG. 12: Jerk parameter vs. z in Riemannian and Finslerian modes.

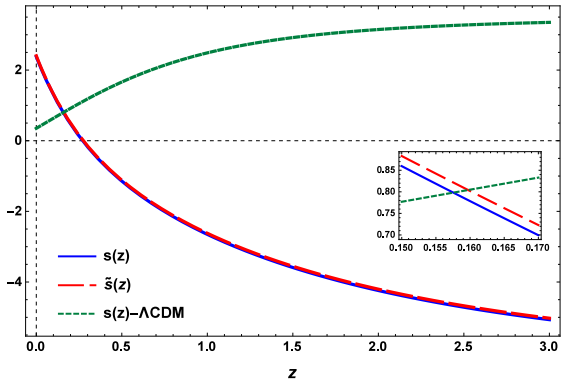


FIG. 13: Snap parameter vs. z in Riemannian and Finslerian modes.

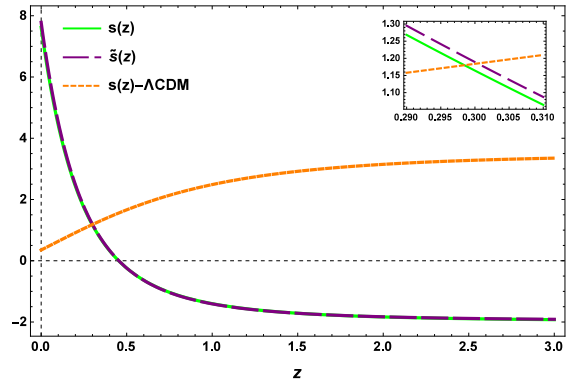


FIG. 14: Snap parameter vs. z in Riemannian and Finslerian modes.

VIII. STATEFINDER DIAGNOSTIC

Statefinder diagnostics [94–97] is generally utilized to denote various dark energy models and compare their behavior using the higher-order derivatives of the scale factor. The parameters are $\mathfrak{s}$ and r are computed by using the following relations:

$$r = j = \frac{\dddot{a}}{aH^3}, \qquad \mathfrak{s} = \frac{r-1}{3\left(q-\frac{1}{2}\right)}.$$

The statefinder diagnostics pairs are made as s, r and q, r wherein various trajectories in the $\mathfrak{s}-r$ and $q-r$ planes scheme to see the temporal evolutions of various dark energy models. The fixed points in this context are generally considered as $\{\mathfrak{s}, r\} = \{0, 1\}$ for the ΛCDM model and $\{\mathfrak{s}, r\} = \{1, 1\}$ for the SCDM (standard cold dark matter) model in FLRW background and the departures of any dark energy model from these fixed points are examined. The other diagnostic pair is $\{q, r\}$ and the fixed points considered are $\{q, r\} = \{-1, 1\}$ for the ΛCDM model and $\{q, r\} = \{0.5, 1\}$ for SCDM model.The statefinder parameters for the considered Model 1 and Model 2 are computed as follows

Model 1

$$\tilde{r} = \tilde{j}, \qquad \tilde{\mathfrak{s}} = \frac{8\alpha\triangle_1}{27\triangle_2}, \tag{30}$$

where

$$\begin{aligned}
\triangle_1 \;=\; & -\Bigg[\left(\frac{3}{2}\,\alpha+\frac{27}{8}\right)(\beta\,(1+z))^{2\,\alpha}+\left(\frac{5}{4}\,\alpha+\frac{15}{8}\right)(\beta\,(1+z))^{3\,\alpha}+\left(\frac{5}{4}\,\alpha-\frac{15}{8}\right)(\beta\,(1+z))^{4\,\alpha}\\
& +\;\left(\frac{5}{4}\,\alpha-\frac{27}{8}\right)(\beta\,(1+z))^{5\,\alpha}+\left(\alpha-\frac{15}{8}\right)(\beta\,(1+z))^{6\,\alpha}+\left(\frac{\alpha}{4}-\frac{3}{8}\right)(\beta\,(1+z))^{7\,\alpha}\\
& +\;\left(\alpha+\frac{15}{8}\right)(\beta\,(1+z))^{\alpha}+\frac{\alpha}{4}+\frac{3}{8}\Bigg]H_0^2\mathrm{e}^{\frac{2\,\beta^{-\alpha}+4+2\,\beta^{\alpha}}{\alpha\,H_0\,(1+(\beta\,(1+z))^{\alpha})}}+\Bigg[\left(\frac{\alpha}{2}+\frac{9}{4}\right)(\beta\,(1+z))^{2\,\alpha}\\
& +\;\left(\frac{\alpha}{2}-\frac{9}{4}\right)(\beta\,(1+z))^{3\,\alpha}+\left(\alpha-\frac{27}{8}\right)(\beta\,(1+z))^{4\,\alpha}+\left(\frac{\alpha}{2}-\frac{9}{8}\right)(\beta\,(1+z))^{5\,\alpha}\\
& +\;\left(\alpha+\frac{27}{8}\right)(\beta\,(1+z))^{\alpha}+\frac{\alpha}{2}+\frac{9}{8}\Bigg](1+z)^{\alpha}\,H_0\left(\frac{1}{2}\,\beta^{2\,\alpha}+\beta^{\alpha}+\frac{1}{2}\right)\mathrm{e}^{\frac{\beta^{-\alpha}+2+\beta^{\alpha}}{\alpha\,H_0\,(1+(\beta\,(1+z))^{\alpha})}}\\
& -\;\frac{3}{4}\,(1+z)^{2\,\alpha}\left(\beta^{\alpha}+\frac{3}{2}\,\beta^{2\,\alpha}+\beta^{3\,\alpha}+\frac{1}{4}\,\beta^{4\,\alpha}+\frac{1}{4}\right)\Bigg((\beta\,(1+z))^{\alpha}-(\beta\,(1+z))^{2\,\alpha}\\
& -\;(\beta\,(1+z))^{3\,\alpha}+1\Bigg),
\end{aligned}$$

$$\begin{aligned}
\triangle_2 \;=\; & \Bigg[-H_0\left(\left(\frac{-\alpha}{3}+\frac{1}{2}\right)(\beta\,(1+z))^{2\,\alpha}+(\beta\,(1+z))^{\alpha}+\frac{\alpha}{3}+\frac{1}{2}\right)\mathrm{e}^{\frac{\beta^{-\alpha}+2+\beta^{\alpha}}{\alpha\,H_0\,(1+(\beta\,(1+z))^{\alpha})}}\\
& +\;(1+z)^{\alpha}\left(\frac{1}{2}\,\beta^{2\,\alpha}+\beta^{\alpha}+\frac{1}{2}\right)\Bigg]\Bigg[-\frac{2}{3}\,H_0\left((\beta\,(1+z))^{\alpha}+\frac{1}{2}\,(\beta\,(1+z))^{2\,\alpha}+\frac{1}{2}\right)\\
& .\;\;\mathrm{e}^{\frac{\beta^{-\alpha}+2+\beta^{\alpha}}{\alpha\,H_0\,(1+(\beta\,(1+z))^{\alpha})}}+(1+z)^{\alpha}\left(\frac{1}{2}\,\beta^{2\,\alpha}+\beta^{\alpha}+\frac{1}{2}\right)\Bigg]\left(1+(\beta\,(1+z))^{\alpha}\right)^3.
\end{aligned}$$

Model 2

$$\tilde{r}=\tilde{j},\qquad \tilde{\mathfrak{s}}=\frac{-3\alpha\triangle_3}{2\triangle_4},\tag{31}$$

$$\begin{aligned}
\triangle_3 \;=\; & -\frac{28\,{H_0}^2\left(1+(\beta\,(1+z))^{2\,\alpha}\right)^{\frac{3}{2}}}{9}\Bigg[\left(\alpha+\frac{15}{14}\right)(\beta\,(1+z))^{2\,\alpha}+\left(\frac{3}{7}\,\alpha+\frac{9}{14}\right)(\beta\,(1+z))^{4\,\alpha}\\
& +\;\left(\frac{\alpha}{7}-\frac{3}{14}\right)(\beta\,(1+z))^{6\,\alpha}+\left(\frac{\alpha}{7}-\frac{3}{14}\right)(\beta\,(1+z))^{8\,\alpha}+\frac{4}{7}\,\alpha+\frac{3}{7}\Bigg]{H_0}^2\mathrm{e}^{2\,\frac{\left(1+\beta^{2\,\alpha}\right)^{\frac{3}{2}}\beta^{-2\,\alpha}}{\alpha\,H_0\,\sqrt{1+(\beta\,(1+z))^{2\,\alpha}}}}\\
& +\;\frac{28\left(1+\beta^{2\,\alpha}\right)^{\frac{3}{2}}(1+z)^{2\,\alpha}}{9}\Bigg[\left(\alpha+\frac{45}{28}\right)(\beta\,(1+z))^{2\,\alpha}+\left(\frac{3}{7}\,\alpha+\frac{27}{28}\right)(\beta\,(1+z))^{4\,\alpha}\\
& +\;\left(\frac{\alpha}{7}-\frac{9}{28}\right)(\beta\,(1+z))^{6\,\alpha}+\left(\frac{\alpha}{7}-\frac{9}{28}\right)(\beta\,(1+z))^{8\,\alpha}+\frac{4}{7}\,\alpha+\frac{9}{14}\Bigg]H_0\,\mathrm{e}^{\frac{\left(1+\beta^{2\,\alpha}\right)^{\frac{3}{2}}\beta^{-2\,\alpha}}{\alpha\,H_0\,\sqrt{1+(\beta\,(1+z))^{2\,\alpha}}}}\\
& +\;\left((\beta\,(1+z))^{2\,\alpha}-2\right)\left(1+(\beta\,(1+z))^{2\,\alpha}\right)^{\frac{3}{2}}(1+z)^{4\,\alpha}\left(\beta^{2\,\alpha}+\beta^{4\,\alpha}+\frac{1}{3}\,\beta^{6\,\alpha}+\frac{1}{3}\right),
\end{aligned}$$

$$\begin{aligned}
\triangle_4 \;=\; & \left(1+(\beta\,(1+z))^{2\,\alpha}\right)^{\frac{5}{2}}\Bigg[\left(1+(\beta\,(1+z))^{2\,\alpha}\right)^{\frac{3}{2}}H_0\,\mathrm{e}^{\frac{\left(1+\beta^{2\,\alpha}\right)^{\frac{3}{2}}\beta^{-2\,\alpha}}{\alpha\,H_0\,\sqrt{1+(\beta\,(1+z))^{2\,\alpha}}}}-\frac{3}{2}\left(1+\beta^{2\,\alpha}\right)^{\frac{3}{2}}\\
& .\;\;(1+z)^{2\,\alpha}\Bigg]\Bigg[\left(\left(\alpha-\frac{3}{2}\right)(\beta\,(1+z))^{2\,\alpha}-2\,\alpha-\frac{3}{2}\right)\sqrt{1+(\beta\,(1+z))^{2\,\alpha}}H_0\\
& .\;\;\mathrm{e}^{\frac{\left(1+\beta^{2\,\alpha}\right)^{\frac{3}{2}}\beta^{-2\,\alpha}}{\alpha\,H_0\,\sqrt{1+(\beta\,(1+z))^{2\,\alpha}}}}+\frac{3}{2}\left(1+\beta^{2\,\alpha}\right)^{\frac{3}{2}}(1+z)^{2\,\alpha}\Bigg].
\end{aligned}$$

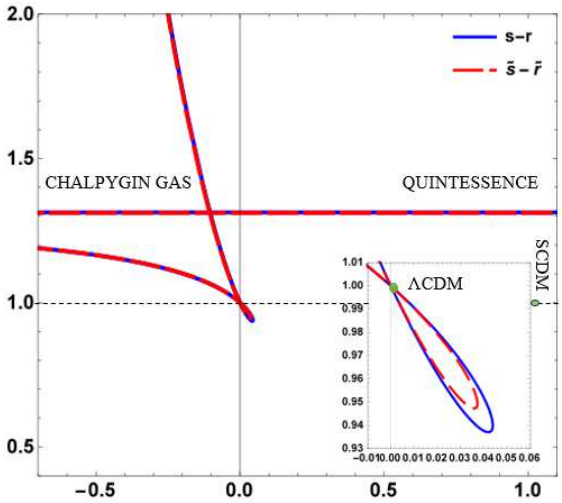


FIG. 15: Model 1: The $\tilde{\mathfrak{s}} - \tilde{r}$ plots show the different trajectories of the models in Riemannian and Finslerian modes.

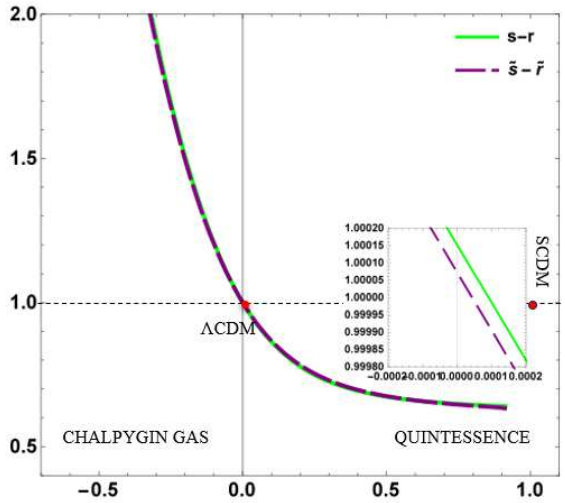


FIG. 16: Model 2: The $\tilde{\mathfrak{s}} - \tilde{r}$ plots show the different trajectories of the models in Riemannian and Finslerian modes.

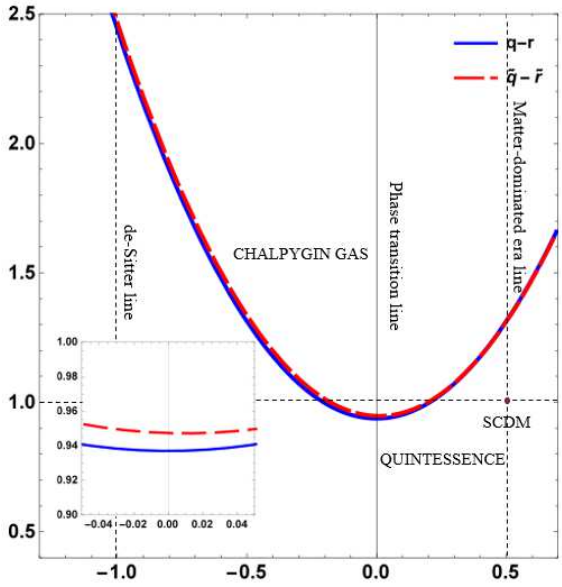


FIG. 17: Model 1: The $\tilde{\mathfrak{q}} - \tilde{r}$ plots show the different trajectories of the models in Riemannian and Finslerian modes.

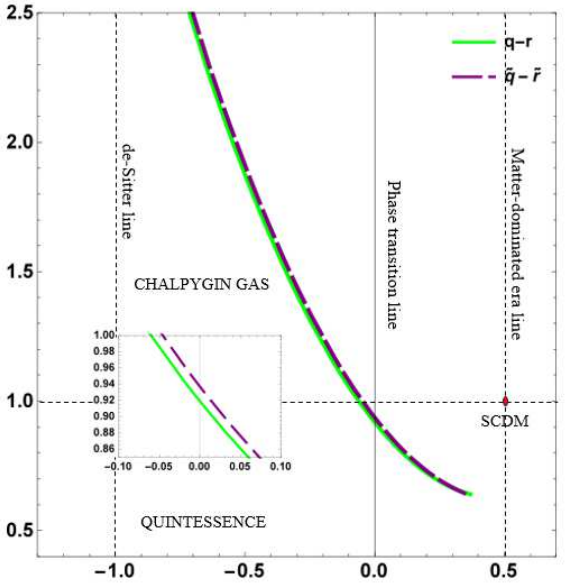


FIG. 18: Model 2: The $\tilde{\mathfrak{q}} - \tilde{r}$ plots show the different trajectories of the models in Riemannian and Finslerian modes.

IX. $Om(z)$ DIAGNOSTIC

Om diagnostic is another instrument defined in [98–101] that uses the Hubble parameter and provides a null test of the ΛCDM model. Similar to the statefinder diagnostic, the $\tilde{O}m$ diagnostic is also a productive method to distinguish different DE models from the ΛCDM model according to the slope variation of $\tilde{O}m(z)$. The positive slope of the diagnostic indicates a quintessence nature ($\omega > -1$), the Negative slope of the diagnostic indicates a phantom nature ($\omega < -1$), and the constant slope with respect to redshift reveals the nature of dark energy coincides with the cosmological constant ($\omega = -1$). The $\tilde{O}m(z)$ for a flat universe is defined as follows:

$$\tilde{Om}(z)=\frac{\left(\frac{\tilde{H}(z)}{H_0}\right)^2-1}{(1+z)^3-1},$$

Model 1

$$\tilde{Om}(z)=\frac{\left(\frac{(1+[\beta(1+z)]^{\alpha})^4}{(1+\beta^{\alpha})^4(1+z)^{2\alpha}}-\frac{(1+[\beta(1+z)]^{\alpha})^2}{H_0(1+\beta^{\alpha})^2(1+z)^{2\alpha}}e^{-\frac{(1+\beta^{\alpha})^2}{\alpha H_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}\right)-1}{(1+z)^3-1},$$

Model 2

$$\tilde{Om}(z)=\frac{\left(\frac{\left(1+[\beta(1+z)]^{2\alpha}\right)^3}{(1+\beta^{2\alpha})^3(1+z)^{4\alpha}}-\frac{\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{3}{2}}}{H_0(1+\beta^{2\alpha})^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}\right)-1}{(1+z)^3-1}.$$

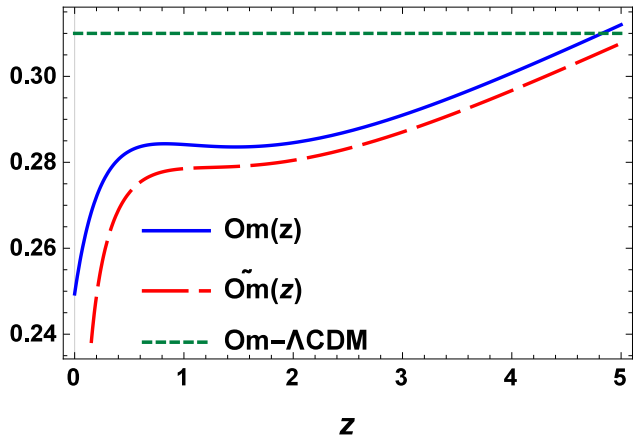


FIG. 19: $Om(z)$ vs. z in Riemannian and Finslerian modes.

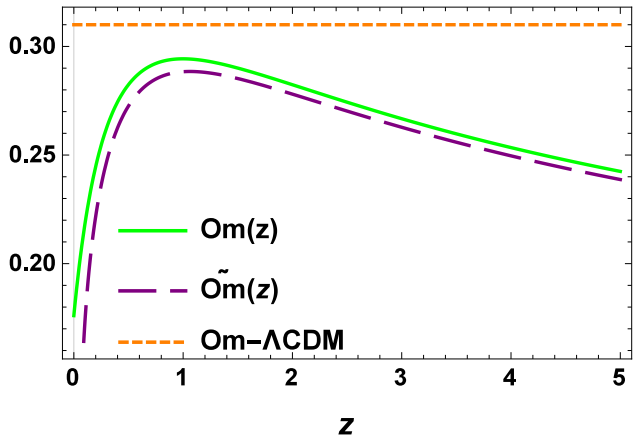


FIG. 20: $Om(z)$ vs. z in Riemannian and Finslerian modes.

X. PHYSICAL PARAMETERS

Here, we consider a two-fluid model, cold dark matter and dark energy, since the radiation contribution at present is negligible. The matter pressure is $P_{\mathfrak{m}}=0$ for cold dark matter, the equation of state is $P_{\mathfrak{de}}=\omega_{\mathfrak{de}}\rho_{\mathfrak{de}}$ for dark energy, and $M_{pl}=(8\pi G)^{-\frac{1}{2}}$. By utilizing Eqs (21) and (22), we derive expressions for critical physical parameters. These expressions allow us to interpret the physical behavior of matter and dark energy densities and pressures. Graphical representations further elucidate these behaviors for both Model 1 and Model 2, providing insights into their respective dynamics within the cosmological framework.

Model 1:

$$\frac{\tilde{P}_{\mathfrak{de}}}{M_{pl}^2H_0^2}=\frac{(-3+2\alpha)\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^4-4\alpha\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^3}{\left(1+\beta^{\alpha}\right)^4(1+z)^{2\alpha}}+5\frac{\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^2}{H_0\left(1+\beta^{\alpha}\right)^2(1+z)^{\alpha}}$$
$$\cdot\ e^{\frac{-(1+\beta^{\alpha})^2}{\alpha H_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}},\qquad(32)$$

$$\frac{\tilde{\rho}_{\mathfrak{de}}}{M_{pl}^2H_0^2} = \frac{1}{\omega_{\mathfrak{de}}}\Bigg\{\frac{(-3+2\alpha)\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^4-4\alpha\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^3}{\left(1+\beta^{\alpha}\right)^4(1+z)^{2\alpha}}+5\frac{\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^2}{H_0\left(1+\beta^{\alpha}\right)^2(1+z)^{\alpha}} . \ \mathrm{e}^{\frac{-(1+\beta^{\alpha})^2}{\alpha \mathrm{H}_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}\Bigg\}, \tag{33}$$

$$\frac{\tilde{\rho}_{\mathfrak{m}}}{M_{pl}^2H_0^2} = \left(3\frac{\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^4}{\left(1+\beta^{\alpha}\right)^4(1+z)^{2\alpha}}-\frac{1}{\omega_{\mathfrak{de}}}\frac{(-3+2\alpha)\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^4-4\alpha\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^3}{\left(1+\beta^{\alpha}\right)^4(1+z)^{2\alpha}}\right) - (3+\frac{5}{\omega_{\mathfrak{de}}})\frac{\left(1+\beta^{\alpha}(1+z)^{\alpha}\right)^2}{H_0\left(1+\beta^{\alpha}\right)^2(1+z)^{\alpha}}\mathrm{e}^{\frac{-(1+\beta^{\alpha})^2}{\alpha \mathrm{H}_0\beta^{\alpha}(1+[\beta(1+z)]^{\alpha})}}. \tag{34}$$

Model 2:

$$\frac{\tilde{P}_{\mathfrak{de}}}{M_{pl}^2H_0^2} = \frac{(-3+2\alpha)\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3-6\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^2}{\left(1+\beta^{2\alpha}\right)^3(1+z)^{4\alpha}}+5\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^{\frac{3}{2}}}{H_0\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}} . \ \mathrm{e}^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha \mathrm{H}_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}, \tag{35}$$

$$\frac{\tilde{\rho}_{\mathfrak{de}}}{M_{pl}^2H_0^2} = \frac{1}{\omega_{\mathfrak{de}}}\Bigg\{\frac{(-3+2\alpha)\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3-6\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^2}{\left(1+\beta^{2\alpha}\right)^3(1+z)^{4\alpha}}+5\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^{\frac{3}{2}}}{H_0\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}} . \ \mathrm{e}^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha \mathrm{H}_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}\Bigg\}, \tag{36}$$

$$\frac{\tilde{\rho}_{\mathfrak{m}}}{M_{pl}^2H_0^2} = \left(3\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3}{\left(1+\beta^{2\alpha}\right)^3(1+z)^{4\alpha}}-\frac{1}{\omega\mathfrak{de}}\frac{(-3+2\alpha)\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3-6\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^2}{\left(1+\beta^{2\alpha}\right)^3(1+z)^{4\alpha}}\right) - (3+\frac{5}{\omega_{\mathfrak{de}}})\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^{\frac{3}{2}}}{H_0\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}. \tag{37}$$

When the candidate of dark energy is considered as cosmological constant, we have $\rho_{\mathfrak{de}} = \rho_{\Lambda}$ and $\omega_{\mathfrak{de}} = -1$, then the evolution of the physical parameters can be illustrated as shown in the following figures: **Scalar Field** Since the equation of state for the cosmological constant is non-dynamical, the observations favor the dynamical equation of state. This feature makes another suitable candidate for depicting dark energy, such as the general scalar field. For a general scalar field ϕ, the action can be denoted as

$$S = \frac{1}{16\pi}\int d^4x\sqrt{-g}\left\{\frac{M_{pl}^2}{2}R-\frac{1}{2}\partial_{\nu}\partial^{\nu}\phi-V(\phi)+{}^{Matter}\mathcal{L}\right\}.$$

In 2018, S. I. Vacaru [102] derived respective energy-momentum tensors for Lagrange-Finsler. Due to this

$$S = {}^{g}S+{}^{\phi}S+{}^{Matter}S, \quad T_{\alpha\beta} = {}^{g}T_{\alpha\beta}+{}^{\phi}T_{\alpha\beta}.$$

In GR, the Lagrange density for gravitational fields is postulated in the form

$${}^{g}\mathcal{L}(g_{ij},\nabla) = \frac{M_{pl}^2}{2}R,$$

where R is the Ricci scalar and the Planck mass M_{pl} is determined by the Newton constant G. $R_{ij} = \frac{8\pi G}{c^4}\left({}^{g}T_{ij}-\frac{1}{2}g_{ij}{}^{g}T\right)$. The Lagrange density for scalar fields is defined in the following form

$${}^{\phi}\mathcal{L} = \frac{1}{2}\dot{\phi}^2-V(\phi),$$

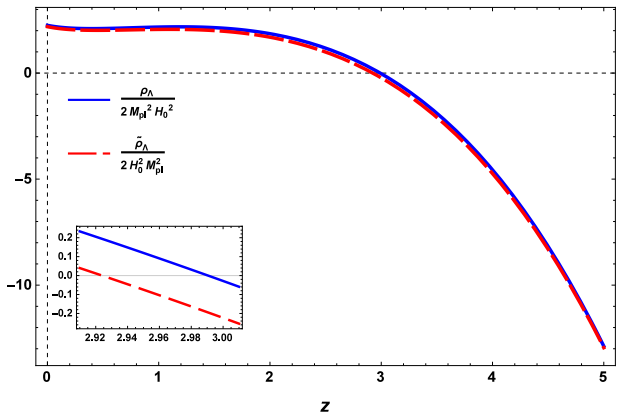


FIG. 21: Model 1: The evolution of the energy densities of the cosmological constant ($\rho_{\mathfrak{de}} = \rho_\Lambda$ and $\omega_{\mathfrak{de}} = -1$) in Riemannian and Finslerian modes.

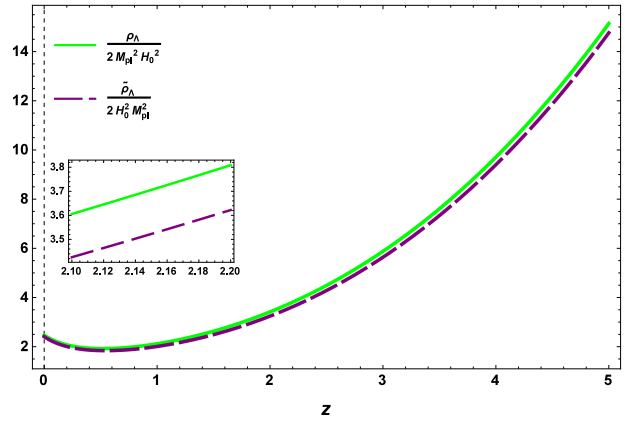


FIG. 22: Model 2: The evolution of the energy densities of the cosmological constant ($\rho_{\mathfrak{de}} = \rho_\Lambda$ and $\omega_{\mathfrak{de}} = -1$) in Riemannian and Finslerian modes.

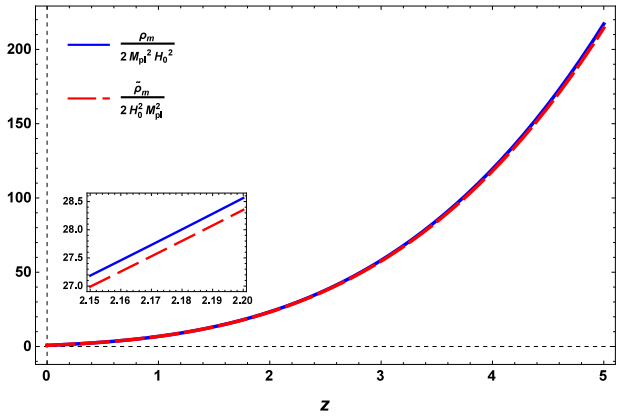


FIG. 23: Model 1: The evolution of the matter energy densities (ρ_m) in Riemannian and Finslerian modes.

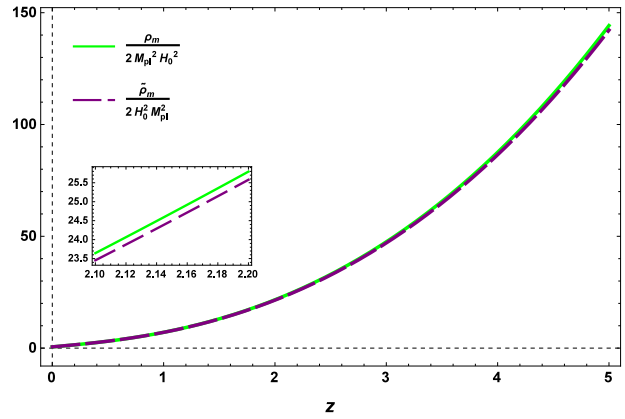


FIG. 24: Model 2: The evolution of the matter energy densities (ρ_m) in Riemannian and Finslerian modes.

where ${}^{\phi}T_{\alpha\beta} = \frac{-2}{\sqrt{|g_{\alpha\beta}|}}\frac{\delta(\sqrt{|g_{\mu\nu}|}{}^{\phi}\mathcal{L})}{\delta g^{\alpha\beta}}$. Therefore, similar to the Riemannian state, we have

$$\tilde{\rho}_\phi = {}^{\phi}T_{00} = \frac{1}{2}\dot{\phi}^2 + V(\phi).$$

For a two-component universe, scalar field and cold dark matter with minimal interaction between them (i.e., $\rho_{\mathfrak{tot}} = \rho_\phi + \rho_{\mathfrak{m}}, \rho_{\mathfrak{m}} = ca^{-3} = c(1+z)^3$, c is a constant of integration), then the solutions obtained from Eqs. (21) and (22) are

Model 1:

$$\begin{aligned}\frac{\tilde{\rho}_\phi}{M_{pl}^2H_0^2} &= \frac{c(1+z)^3}{M_{pl}^2H_0^2} - \left(\frac{2\alpha\left(1+\beta^\alpha(1+z)^\alpha\right)^4 - 4\alpha\left(1+\beta^\alpha(1+z)^\alpha\right)^3}{\left(1+\beta^\alpha\right)^4(1+z)^{2\alpha}}\right)\\ &- 2\frac{\left(1+\beta^\alpha(1+z)^\alpha\right)^2}{H_0\left(1+\beta^\alpha\right)^2(1+z)^\alpha}\mathrm{e}^{\frac{-(1+\beta^\alpha)^2}{\alpha H_0\beta^\alpha(1+[\beta(1+z)]^\alpha)}},\end{aligned} \tag{38}$$

$$\frac{\tilde{V}(\phi)}{M_{pl}^2H_0^2} = \frac{c(1+z)^3}{2M_{pl}^2H_0^2} - \left(\frac{(\frac{-3}{2}+2\alpha)\left(1+\beta^\alpha(1+z)^\alpha\right)^4 - 4\alpha\left(1+\beta^\alpha(1+z)^\alpha\right)^3}{(1+\beta^\alpha)^4(1+z)^{2\alpha}}\right) - \frac{7}{2}\frac{(1+\beta^\alpha(1+z)^\alpha)^2}{H_0(1+\beta^\alpha)^2(1+z)^\alpha}e^{\frac{-(1+\beta^\alpha)^2}{\alpha H_0\beta^\alpha(1+[\beta(1+z)]^\alpha)}}. \quad (39)$$

Model 2:

$$\frac{\tilde{\rho}_\phi}{M_{pl}^2H_0^2} = \frac{c(1+z)^3}{M_{pl}^2H_0^2} - \left(\frac{2\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3 - 6\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^2}{(1+\beta^{2\alpha})^3(1+z)^{4\alpha}}\right) - 2\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^{\frac{3}{2}}}{H_0(1+\beta^{2\alpha})^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}, \quad (40)$$

$$\frac{\tilde{V}(\phi)}{M_{pl}^2H_0^2} = \frac{c(1+z)^3}{2M_{pl}^2H_0^2} - \left(\frac{(\frac{-3}{2}+2\alpha)\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3 - 6\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^2}{(1+\beta^{2\alpha})^3(1+z)^{4\alpha}}\right) - \frac{7}{2}\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^{\frac{3}{2}}}{H_0(1+\beta^{2\alpha})^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}. \quad (41)$$

As we know, the total density parameters is written as follows,

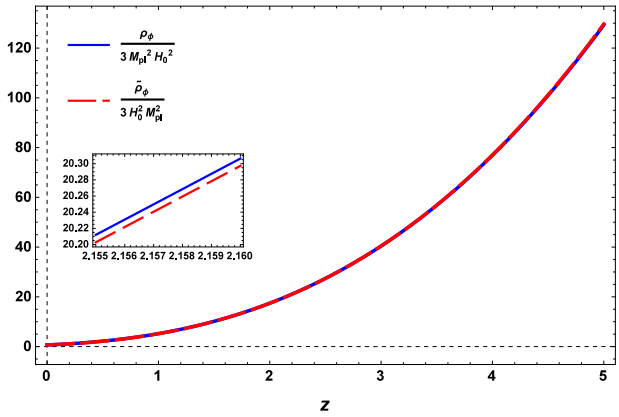


FIG. 25: Model 1: The evolution of the scalar field energy density (ρ_ϕ) in Riemannian and Finslerian modes.

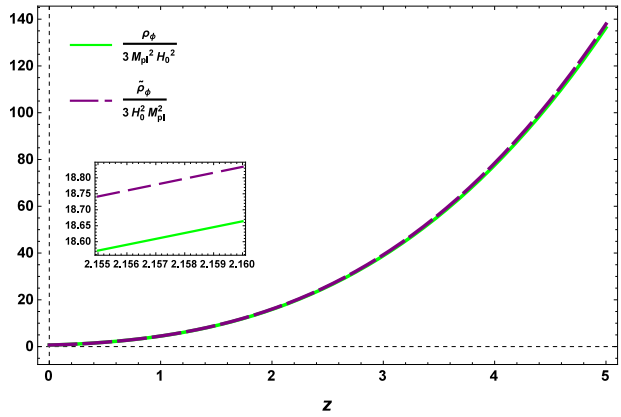


FIG. 26: Model 2: The evolution of the scalar field energy density (ρ_ϕ) in Riemannian and Finslerian modes.

$$1 = \tilde{\Omega} = \tilde{\Omega}_\phi + \tilde{\Omega}_{\mathrm{m}} \quad (42)$$

where $\tilde{\Omega}_\phi = \frac{\tilde{\rho}_\phi}{3M_{pl}^2H^2}$ and $\tilde{\Omega}_{\mathrm{m}} = \frac{\tilde{\rho}_{\mathrm{m}}}{3M_{pl}^2H^2}$ are density parameter for the scalar field and density parameter for the matter, respectively. They can be computed for two models as,

Model 1:

$$\tilde{\Omega}_\phi = \frac{\frac{c(1+z)^3}{M_{pl}^2H_0^2} - \left(\frac{2\alpha(1+\beta^\alpha(1+z)^\alpha)^4 - 4\alpha(1+\beta^\alpha(1+z)^\alpha)^3}{(1+\beta^\alpha)^4(1+z)^{2\alpha}}\right) - 2\frac{(1+\beta^\alpha(1+z)^\alpha)^2}{H_0(1+\beta^\alpha)^2(1+z)^\alpha}e^{\frac{-(1+\beta^\alpha)^2}{\alpha H_0\beta^\alpha(1+[\beta(1+z)]^\alpha)}}}{3\left(\frac{(1+[\beta(1+z)]^\alpha)^4}{(1+\beta^\alpha)^4(1+z)^{2\alpha}} - \frac{(1+[\beta(1+z)]^\alpha)^2}{H_0(1+\beta^\alpha)^2(1+z)^{2\alpha}}e^{-\frac{(1+\beta^\alpha)^2}{\alpha H_0\beta^\alpha(1+[\beta(1+z)]^\alpha)}}\right)}, \quad (43)$$

from Eqs. (42) and (43), we obtain

$$\frac{c}{M_{pl}^2H_0^2} = 3(1-\Omega_{\mathrm{m}_0})(1-\frac{1}{H_0}e^{\frac{-(1+\beta^\alpha)}{\alpha H_0\beta^\alpha}}) + (2\alpha - \frac{4\alpha}{1+\beta^\alpha}) + \frac{2}{\mathrm{H}_0}e^{\frac{-(1+\beta^\alpha)}{\alpha H_0\beta^\alpha}} = \mathcal{A}, \quad (44)$$

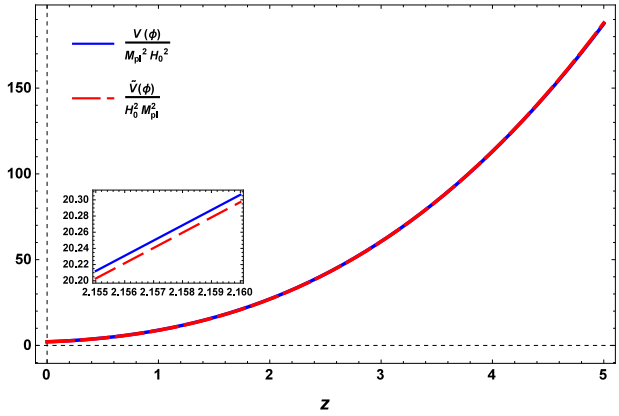


FIG. 27: Model 1: The evolution of the scalar field potential ($V(\phi)$) in Riemannian and Finslerian modes.

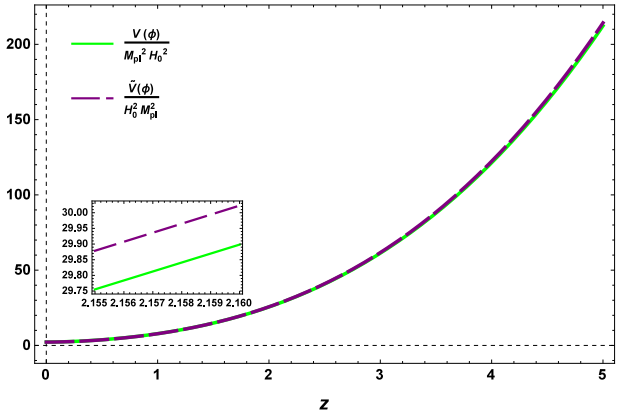


FIG. 28: Model 2: The evolution of the scalar field potential ($V(\phi)$) in Riemannian and Finslerian modes.

the equations of state parameter ($\tilde{\omega}_\phi = \frac{\tilde{P}_\phi}{\tilde{\rho}_\phi}$) are given by

$$\tilde{\omega}_\phi = \frac{\frac{(-3+2\alpha)(1+\beta^\alpha(1+z)^\alpha)^4-4\alpha(1+\beta^\alpha(1+z)^\alpha)^3}{(1+\beta^\alpha)^4(1+z)^{2\alpha}} + 5\frac{(1+\beta^\alpha(1+z)^\alpha)^2}{H_0(1+\beta^\alpha)^2(1+z)^\alpha}e^{\frac{-(1+\beta^\alpha)^2}{\alpha H_0\beta^\alpha(1+[\beta(1+z)]^\alpha)}}}{\mathcal{A}(1+z)^3 - \left(\frac{2\alpha(1+\beta^\alpha(1+z)^\alpha)^4-4\alpha(1+\beta^\alpha(1+z)^\alpha)^3}{(1+\beta^\alpha)^4(1+z)^{2\alpha}}\right) - 2\frac{(1+\beta^\alpha(1+z)^\alpha)^2}{H_0(1+\beta^\alpha)^2(1+z)^\alpha}e^{\frac{-(1+\beta^\alpha)^2}{\alpha H_0\beta^\alpha(1+[\beta(1+z)]^\alpha)}}}. \tag{45}$$

Model 2:

$$\tilde{\Omega}_\phi = \frac{\frac{c(1+z)^3}{M_{pl}^2H_0^2} - \left(\frac{2\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3-6\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^2}{(1+\beta^{2\alpha})^3(1+z)^{4\alpha}}\right) - 2\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^{\frac{3}{2}}}{H_0(1+\beta^{2\alpha})^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}}{3\left(\frac{\left(1+[\beta(1+z)]^{2\alpha}\right)^3}{(1+\beta^{2\alpha})^3(1+z)^{4\alpha}} - \frac{\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{3}{2}}}{H_0(1+\beta^{2\alpha})^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}\right)}, \tag{46}$$

from Eqs. (42) and (46), we obtain

$$\frac{c}{M_{pl}^2H_0^2} = 3(1-\Omega_{\mathfrak{m}_0})(1-\frac{1}{H_0}e^{\frac{-(1+\beta^{2\alpha})}{\alpha H_0\beta^{2\alpha}}}) + (2\alpha - \frac{6\alpha}{1+\beta^{2\alpha}}) + \frac{2}{H_0}e^{\frac{-(1+\beta^{2\alpha})}{\alpha H_0\beta^{2\alpha}}} = \mathcal{A}, \tag{47}$$

the equations of state parameter ($\tilde{\omega}_\phi = \frac{\tilde{P}_\phi}{\tilde{\rho}_\phi}$) are given by

$$\tilde{\omega}_\phi = \frac{\frac{(-3+2\alpha)\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^3-6\alpha\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^2}{(1+\beta^{2\alpha})^3(1+z)^{4\alpha}} + 5\frac{\left(1+\beta^{2\alpha}(1+z)^{2\alpha}\right)^{\frac{3}{2}}}{H_0(1+\beta^{2\alpha})^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{\left(1+\beta^{2\alpha}\right)^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}\left(1+[\beta(1+z)]^{2\alpha}\right)^{\frac{1}{2}}}}}{\mathcal{A}(1+z)^3 - \left(\frac{2\alpha(1+\beta^{2\alpha}(1+z)^{2\alpha})^3-6\alpha(1+\beta^{2\alpha}(1+z)^{2\alpha})^2}{(1+\beta^{2\alpha})^3(1+z)^{4\alpha}}\right) - 2\frac{(1+\beta^{2\alpha}(1+z)^{2\alpha})^{\frac{3}{2}}}{H_0(1+\beta^{2\alpha})^{\frac{3}{2}}(1+z)^{2\alpha}}e^{-\frac{(1+\beta^{2\alpha})^{\frac{3}{2}}}{\alpha H_0\beta^{2\alpha}(1+[\beta(1+z)]^{2\alpha})^{\frac{1}{2}}}}}. \tag{48}$$

XI. STATISTICAL ANALYSIS

When comparing the dark energy models in both Riemannian and Finslerian modes with the ΛCDM model, it's important to recognize that the ΛCDM model is encompassed within both dark energy models. This nesting property allows for the application of standard statistical tests. We compute these tests using the complete dataset. Consequently, one can define the reduced chi-square statistic as $\chi^2_{\rm red} = \frac{\chi^2}{\rm DoF}$, where DoF represents the degrees of

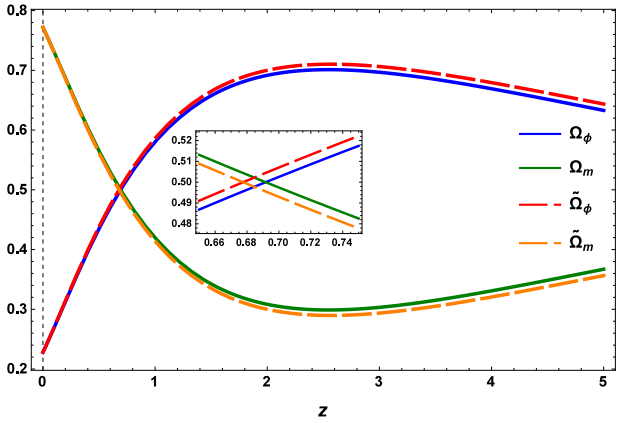


FIG. 29: Model 1: The evolution of the density parameters Ω_ϕ and Ω_m in Riemannian and Finslerian modes.

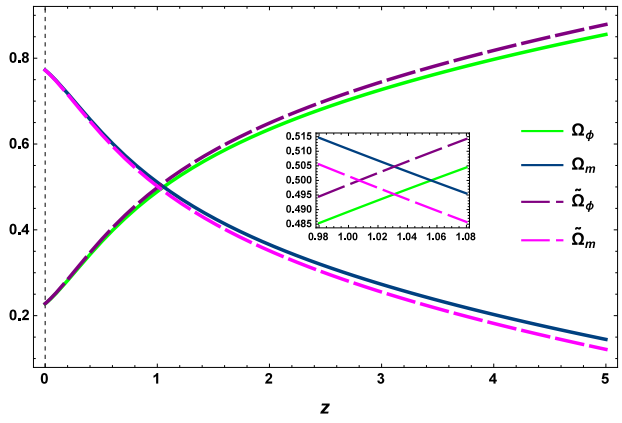


FIG. 30: Model 2: The evolution of the density parameters Ω_ϕ and Ω_m in Riemannian and Finslerian modes.

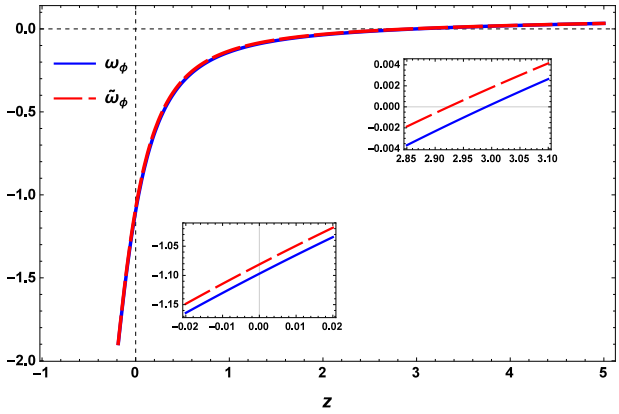


FIG. 31: Model 1: The evolution of equation of state parameter ω_ϕ in Riemannian and Finslerian modes.

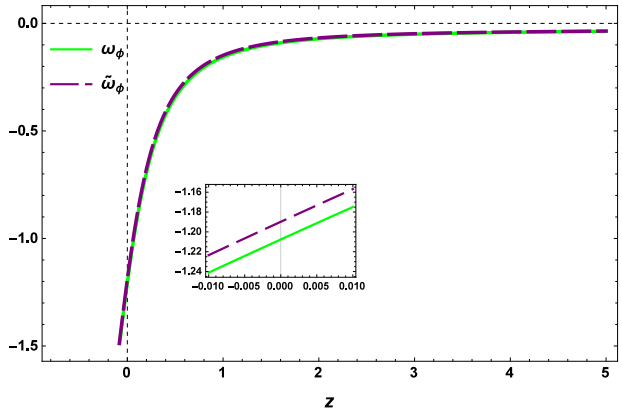


FIG. 32: Model 2: The evolution of equation of state parameter ω_ϕ in Riemannian and Finslerian modes.

freedom of the model, and χ^2 denotes the weighted sum of squared deviations. This calculation is performed under the constraint of the same number of runs. Certainly, here's an addition to the text that includes the P test [103–105]: The Akaike information criterion (AIC) [106, 107] provides a means of comparing different models, considering a large dataset (in our case $N = 273$). It is defined as follows: $\mathrm{AIC} = -2\ln(\mathcal{L}_{\max}) + 2k$. We utilize the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) to evaluate our cosmological models. Specifically, the AIC is defined as:

$$\mathrm{AIC} = -2\ln(\mathcal{L}_{\max}) + 2k + \frac{2k(2k+1)}{N_{\mathrm{tot}} - k - 1}, \tag{49}$$

where $\mathcal{L}_{\max}$ represents the maximum likelihood of the considered data, N_{tot} is the total number of data points, and k denotes the number of parameters. For large N_{tot}, the expression simplifies to:

$$\mathrm{AIC} \approx -2\ln(\mathcal{L}_{\max}) + 2k, \tag{50}$$

which is the standard form of the AIC criterion. Conversely, the Bayesian information criterion is defined as:

$$\mathrm{BIC} = -2\ln(\mathcal{L}_{\max}) + k\ln N_{\mathrm{tot}}. \tag{51}$$

In addition to AIC and BIC, we can also consider the P-test, which evaluates the significance of the difference between the two models. This test quantifies whether the improvement in fit provided by a more complex model (compared to a simpler one) is statistically significant. It is commonly used in model selection to determine whether additional parameters introduced in a more intricate model are justified by improved fit. The P-test complements AIC and BIC in providing model performance and complexity with a comprehensive assessment.

Model	$\mathcal{L}_{\max}$	χ^2_{red}	AIC	BIC	P-Value
ΛCDM Model	254.53	0.9497	264.53	264.76	0.7127
Model 1 Finslerian Geometry	253.32	0.9452	263.32	263.55	0.7307
Model 1 Riemannian Geometry	253.60	0.9462	263.60	263.82	0.7267
Model 2 Finslerian Geometry	250.75	0.9356	260.68	260.90	0.7675
Model 2 Riemannian Geometry	250.68	0.9353	260.75	260.98	0.7686

TABLE VII: Summary of the $\mathcal{L}_{\max}$, χ^2_{red}, AIC, BIC and P-Value

XII. RESULTS AND DISCUSSION

a. Deceleration parameter Figs. 9 and 10 illustrate the evolution of the deceleration parameter (q) in terms of redshift for both Model 1 and Model 2, in both Riemannian and Finslerian geometries, compared to the standard ΛCDM model. In Fig. 9, one can observe that in the distant past ($z \to \infty$), Model 1 predicts values of approximately 0.6932 in Finslerian geometry and 0.6921 in Riemannian geometry. In contrast, the ΛCDM model predicts a value around 0.449. This behavior suggests that in the early Universe, both Model 1 variants exhibit higher deceleration compared to the ΛCDM model. As the redshift approaches zero ($z \to 0$), the present-day deceleration parameter (q_0) is approximately -0.6238 in Finslerian geometry and -0.6227 in Riemannian geometry. In contrast, the ΛCDM model predicts a value around -0.5506. This indicates that, in the current epoch, both geometrical models (Finslerian and Riemannian) predict a slightly higher Universe acceleration compared to the ΛCDM model. The phase transition redshift (z_{tr}) is 0.6958 in Finslerian geometry, 0.7066 in Riemannian geometry, and 0.6824 in the ΛCDM model. The phase transition redshift marks the point at which the Universe transitions from decelerated to accelerated expansion. In Finslerian and Riemannian geometries, this transition occurs slightly later than in the ΛCDM model. Fig. 10 illustrates the behavior of Model 2 in Finslerian geometry, contrasting it with the ΛCDM model. In the distant past ($z \to \infty$), Model 2 predicts deceleration parameter values of approximately 0.3656 in Finslerian geometry and 0.3643 in Riemannian geometry, compared to the ΛCDM model's 0.4490, indicating higher deceleration in both geometrical models. At present ($z \to 0$), the deceleration parameter (q_0) is approximately -0.7352 in Finslerian geometry and -0.7323 in Riemannian geometry. While the ΛCDM model predicts -0.5506, suggesting a slightly higher current acceleration in both geometrical models. The phase transition redshift (z_{tr}), marking the shift from decelerated to accelerated expansion, is 0.6072 in Finslerian geometry, 0.6176 in Riemannian geometry, and 0.6824 in the ΛCDM model, indicating a slightly later transition in the alternative geometries. In each case, the Universe initially experiences a smooth deceleration phase, indicating that no inflationary phase exists included in these models. As redshift decreases, the Universe transitions from this decelerating expansion to an accelerating expansion phase. It can be observed in the figures that the phase transition occurs earlier in the Finslerian mode compared to the Riemannian mode.

b. Jerk parameter Figs. 11 and 12 illustrate the evolution of the jerk parameter (j) in terms of redshift. At high redshifts, Model 1 and Model 2, in both Riemannian and Finslerian geometries, exhibit deviations from the ΛCDM model. This deviation is evident as the jerk parameter value of the ΛCDM model remains constant at $j = 1$ across all redshifts. At redshift z_0, for Model 1 in Finslerian geometry, the jerk parameter value is 1.5465 times that predicted by the ΛCDM model. Similarly, in Riemannian geometry, the jerk parameter value for Model 1 is also 1.5247 times that predicted by the ΛCDM model. For Model 2, in Finslerian geometry, the predicted value j_0 is 2.5956 times that of the ΛCDM model, while in Riemannian geometry, j_0 is 2.557 times the value predicted by the ΛCDM model. The deviation of jerk parameter values from the ΛCDM model suggests that both Model 1 and Model 2 exhibit different evolutionary behaviors, particularly at high redshifts. The higher values of the jerk parameter at z_0 compared to the ΛCDM model imply that these alternative models may incorporate additional factors or mechanisms influencing cosmic acceleration, indicating potential differences in the underlying dynamics of cosmic expansion.

c. Snap parameter Figs. 13 and 14 illustrate the evolution of the snap parameter (s) with respect to redshift. At high redshifts, Model 1 and Model 2, under both Riemannian and Finslerian geometries, deviate from the ΛCDM model. Specifically, Model 1 and Model 2 predict negative snap values, whereas the ΛCDM model predicts positive values of snap. At redshift z_0, for Model 1 in both the Finslerian and Riemannian modes, the predicted values are 2.4039 and 2.3807, respectively. In the case of Model 2, in both the Finslerian and Riemannian modes, the predicted values are 7.7836 and 7.6217, respectively, while the ΛCDM model predicts 0.3573. The negative values of the snap

parameter predicted by Model 1 and Model 2 suggest that these models exhibit behaviors not captured by the ΛCDM model, particularly at high redshifts. Additionally, the significantly higher values of snap at present (z_0) in both models compared to the ΛCDM model indicate that these alternative models may involve additional physical effects or cosmological components that contribute to the observed cosmic dynamics.

d. Statefinder diagnostic In Figs. 15 and 16, one can see the divergence evolutions of Model 1 and Model 2 in the $\mathfrak{s} - r$ planes. Both models show distinct characteristics as compared to the other standard models. One can observe that at early times, Model 1 presumes values in the range $r > 1$ and $\mathfrak{s} < 0$ denoting Chaplygin gas type DE model and evolutes to the quintessence region and again retreats to Chapligyn gas region at late times by crossing the intermediate ΛCDM fixed point $\{0, 1\}$ during evolution. But, Model 2 is different, and evolutes from the quintessence region in the past and moves to the Chaplygin gas region intermediating, the ΛCDM fixed point $\{0, 1\}$ while evolving for all cases. Figs. 17 and 18 represent the temporal evolution of Model 1 and Model 2 in the $\{q, r\}$ plane and provide additional information about both models wherein the dashed lines describe the evolution of the ΛCDM model below which quintessence region and the upper one is Chaplygin gas region are shown. The evolution of Model 1 and Model 2 are observed. Both models deviates from de Sitter point $(-1, 1)$.

e. $Om(z)$ diagnostic Figs. 19 and 20 show the evolution of the $Om(z)$ diagnostic parameter in terms of redshift z for Model 1 and Model 2 in both Riemannian and Finslerian geometries. For Model 1, $Om(z)$ consistently decreases (negative slope) across all redshifts, indicating quintessence-like dynamics. In contrast, for Model 2, $Om(z)$ initially increases (positive slope) for $z > 1$, suggesting a phantom-like evolution, and then decreases (negative slope) at lower redshifts, indicating a return to quintessence-like dynamics.

f. Energy Density & Pressure of DE The cosmological evolution of the energy density of dark energy, indicated by $\rho_{\mathfrak{de}}$, and the dark energy pressure $P_{\mathfrak{de}}$ for our models, Model 1 and Model 2 supply valuable perspicuity into the behavior and effects of dark energy in a cosmological context. In the above plotted figures: Figs. 21, 22, 23, and 24, we have shown the evolution of the dark energy densities and the matter-energy densities when the cosmological constant is considered as a candidate of dark energy. They show almost similar behaviors in the standard mode in both the Riemannian and Finslerian modes for Model 1 and Model 2. Similarly, when a scalar field is considered to be a candidate of dark energy, the evolution of the physical parameters such as the time-dependent dark energy densities, pressures, energy density of cold dark matter, the potential $V(\phi)$ and the equation of state parameter ω_ϕ together with the density parameters Ω_ϕ are shown in the above figures: Figs. 25, 26, 27, 28, 29, 30, 31, and 32 depicting the physical evolution of the Universe for both the models with the solid and dash lines representing the Riemannian and Finslerian modes, respectively. Since we are more concerned about the late Universe, we have kept the redshift range as (0,5) only in most of these figures. We can also predict the values of these physical parameters at redshift $z = 0$. For example, In the case of Model 1, as redshift approaches zero ($z \to 0$), the present-day energy density (ρ_ϕ) is approximately 0.6746 in Finslerian geometry and (ρ_ϕ) assumes value 0.6842 in Riemannian geometry. And, in the case of Model 2, as $z \to 0$, ρ_ϕ is approximately 0.6746 in Finslerian geometry and 0.6830 in Riemannian geometry. As the redshift approaches zero ($z \to 0$), the present-day EoS parameter ω_ϕ is approximately -1.0812 in Finslerian geometry and -1.0971 in Riemannian geometry for Model 1 and for Model 2, ω_ϕ is approximately -1.18981 in Finslerian geometry and -1.2077 in Riemannian geometry. Similarly, we can see the evolution of the other physical parameters as shown in the figures.

g. Statistical analysis Based on the information provided in Table VII, let's conduct a comparative study: The highest likelihood is obtained for the ΛCDM model, followed closely by both Model 1 and Model 2 in both Finslerian and Riemannian geometries. This indicates that all models provide reasonable fits to the data, with slight variations in their maximum likelihood values. The reduced chi-square values are close to 1 for all models, indicating that the models adequately describe the data within the uncertainties. Again, no significant differences are observed between the models in terms of their goodness of fit. Lower values of AIC indicate better model performance. The ΛCDM model has the lowest AIC suggesting that it provides the best balance between goodness of fit and model complexity. However, the differences in AIC between the ΛCDM model and the Finslerian and Riemannian models are relatively small. Similarly, lower BIC values indicate better model performance. Again, the ΛCDM model has the lowest BIC, indicating its preference over the alternative models. However, similar to AIC, the differences in BIC between the ΛCDM model and the Finslerian and Riemannian models are not substantial. The P-values for all models are relatively high, ranging from 0.7127 to 0.7686. These values indicate that the models are statistically consistent

with the observed data, with no strong evidence against them. Overall, while the ΛCDM model performs slightly better according to AIC and BIC, the differences between the Finslerian and Riemannian models and the ΛCDM model are marginal. However, if we consider simplicity and theoretical elegance, Finslerian geometry may be favored over Riemannian geometry. Finslerian geometry offers additional flexibility and generality compared to Riemannian geometry, allowing for a more comprehensive description of the underlying physics, which could lead to a better overall understanding of the universe's dynamics.

XIII. CONCLUSION

In this paper, with the aim of whether the space-time of our universe could be a Finsler-Randers manifold instead of a Riemannian one, we have initiated studying an FRW model with a weak vector field embedded in the metric structure of space-time. The metric structure supplies us with the extended Friedmann equation like 17. The contribution of the variation of anisotropy is described by the extra parameter $\dot{\mathfrak{u}}_0$ yielded by the Finslerian space-time. Specifically, as it is obvious from equation $2\dot{C}_{000} = \dot{\mathfrak{u}}_0$, $\dot{\mathfrak{u}}_0$ directly depends on the Cartan torsion component C_{000}. We considered the Finsler-Randers space-time to offer a novel perspective on cosmic dynamics, departing from the constraints of general relativity. We proposed two dark energy models resulting from the parametrization of H within this geometric framework. We have derived the field equations governing the universe's evolution within the Finsler-Randers formalism, incorporating the presence of dark energy. Through this, we explored its implications on cosmological phenomena, including cosmic expansion, late-time behavior of the universe, cosmological phase transition, and a few more. The comparative analysis of Model 1 and Model 2 in Riemannian and Finslerian geometries against the standard ΛCDM model yields substantial insights into the dynamics of the universe's expansion and the behavior of dark energy. The examination of the Hubble parameter $H(z)$, Hubble difference $\Delta H(z)$, and distance modulus $\mu(z)$ reveals that deviations from the ΛCDM model become significant at higher redshifts. Model 1 shows noticeable deviations for $z > 0.6$, whereas Model 2 exhibits deviations for $z > 1.2$. At lower redshifts, all models demonstrate close agreement, particularly with observational data, underscoring the alignment of the alternative geometries with the ΛCDM model during recent epochs. This close agreement at lower redshifts, where dark energy dominates the expansion dynamics, highlights the robustness of the Riemannian and Finslerian formulations in capturing the late-time evolution of the universe. The analysis of the deceleration parameter (q) indicates that both models predict higher deceleration in the early universe and slightly higher acceleration in the current epoch compared to the ΛCDM model. Additionally, the phase transition redshift (z_{tr}) occurs slightly later in both alternative geometries. The jerk parameter (j) shows that both models predict significantly higher present-day values than the ΛCDM model, suggesting different evolutionary behaviors and potential additional mechanisms influencing cosmic acceleration. The snap parameter (s) also deviates from the ΛCDM model, with both alternative models predicting higher present-day values and negative values at high redshifts, indicating the presence of additional physical effects or cosmological components. The Statefinder diagnostic further distinguishes the evolutionary paths of Model 1 and Model 2 in the $\mathfrak{s} - r$ and $\{q, r\}$ planes, showing transitions through various regions associated with different dark energy models and highlighting their differences from the ΛCDM model. The $Om(z)$ diagnostic suggests that $Om(z)$ decreases for Model 1, indicating quintessence. Model 2 exhibits a more complex behavior with an initial increase (phantom-like) at $z > 1$ followed by a decrease (quintessence-like) at lower redshifts. Statistical analysis using likelihood, reduced chi-square, AIC, and BIC values reveals that while the ΛCDM model performs slightly better overall, the differences between it and the alternative models are marginal. The high P-values indicate that all models are statistically consistent with the observed data. This statistical consistency underscores the viability of the Finslerian and Riemannian models as alternatives to the ΛCDM framework. In conclusion, the ΛCDM model remains slightly preferable based on statistical metrics; however, the Finslerian and Riemannian models offer comparable fits and provide additional flexibility and generality. Finslerian geometry, in particular, may offer a more comprehensive framework for understanding the Universe's dynamics, potentially leading to new insights and a deeper understanding of cosmic evolution. The subtle differences in the evolution of physical parameters between Riemannian and Finslerian geometries, especially at higher redshifts, highlight the potential of Finslerian geometry to accommodate a broader range of cosmological phenomena, making it a promising candidate for future explorations in cosmology.

[1] V. Sahni, A. Starobinsky, The case for a positive cosmological Λ- terma, Int. J. Mod. Phys. D, **9**, 373 (2000).
[2] R. G. Vishwakarma, Consequences for some dark energy candidates from the type la supernova SN 1997ff, Mon. Not. R. Astron. Soc., **331**, 776 (2002).
[3] P. J. E. Peebles and B. Ratra, The cosmological constant and dark energy, Rev. Mod. Phys., **75**, 559 (2003).
[4] S. Perlmutter et al., Measurements of Ω and Λ from 42 High-Redshift Supernovae, Astrophys. J., **517**, 565 (1999).

[5] A. Riess et al., BVRI light curves for 22 type Ia supernovae, Astrophys. J., **117**, 707 (1999).
[6] A. H. Jaffe et al., Cosmology from MAXIMA-1, BOOMERANG, and COBE DMR Cosmic Microwave Background Observations, Phys. Rev. Lett., **86**, 3475 (2001)
[7] D. N. Spergel et al., Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology, Astrophys. J. Suppl., **170**, 377 (2007).
[8] P. de Bernardis et al., A flat Universe from high-resolution maps of the cosmic microwave background radiation, Nature, **404**, 955–959 (2000).
[9] S. Hanany et al., MAXIMA-1: A Measurement of the Cosmic Microwave Background Anisotropy on Angular Scales of $10' - 5^0$, Astrophy. J. Lett., **545**, L5 (2000).
[10] A. G. Riess et al., Tests of the Accelerating Universe with Near-Infrared Observations of a High-Redshift Type Ia Supernova, Astrophy. J., **536**, 62 (2000).
[11] A. G. Riess et al., New Hubble Space Telescope Discoveries of Type Ia Supernovae at $z \geq 1$: Narrowing Constraints on the Early Behavior of Dark Energy*, Astrophy. J., **659**, 98 (2007).
[12] D. H. Weinberg et al., Observational probes of cosmic acceleration, Phys. Rept. **530**, 87 (2013).
[13] P. Astier et al., The Supernova Legacy Survey: measurement of , and w from the first year data set, Astron. & Astrophy., **447**, 31 (2006).
[14] Amanullah, et al., Spectra and Hubble Space Telescope Light Curves of Six Type Ia Supernovae at $0.511 < z < 1.12$ and the Union2 Compilation, Astrophy. J., **716**, 712–738 (2010).
[15] R. G. Vishwakarma, A Machian Approach to General Relativity, Int. J. Geom. Meth. Mod. Phys., **12**, 1550116 (2015).
[16] R. G. Vishwakarma, Mysteries of $R^{ik} = 0$: A novel paradigm in Einstein's theory of gravitation, Front. Phys. **9**, 98 (2014).
[17] R. G. Vishwakarma, Mysteries of the geometrization of gravitation, Res. Astron. Astrophys., **13**, 1409 (2013).
[18] K. Schwarzschild, On the gravitational field of a mass point according to Einstein's theory, Sitzungsb. der Konig. Preuss. Akad. der Wissen., **7**, 189 (1916).
[19] R. C. Tolman, Static Solutions of Einstein's Field Equations for Spheres of Fluid, Phys. Rev., **55**, 364 (1939).
[20] W. de Sitter, Einstein's theory of gravitation and its astronomical consequences, Mon. Not. R. Astron. Soc., **76**, 699 (1916).
[21] R. J. Adlar, A Fluid Sphere in General Relativity, J. Math. Phys., **15**, 727 (1974).
[22] P. C. Vaidya, R. Tikekar, Exact relativistic model for a superdense star, J. Astrophys. & Astron., **3**, 325 (1982).
[23] H. Knutsen, Some physical properties and stability of an exact model of a relativistic star, Astrophys. & Space Science, **140**, 385 (1988).
[24] H. A. Buchdahl, General-Relativistic Fluid Spheres. III. a Static Gaseous Model, Astrophys. J., **147**, 310 (1967).
[25] M. C. Durgapal, A Class of New Exact Solutions in General Relativity. Journal of Physics A: Mathematical and General, J. Phys. A. Math. Gen., **15**, 2637 (1982).
[26] S. K. J. Pacif, Dark energy models from a parametrization of H: a comprehensive analysis and observational constraints, Eur. Phys. J. Plus, **135**, 792 (2020).
[27] J. K. Singh, R. Nagpal, FLRW cosmology with EDSFD parametrization, Eur. Phys. J. C, **80**, 295 (2020).
[28] A. P. Kale, Y. S. Solanke, S. H. Shekh, A. Pradhan Transit f (Q, T) Gravity Model: Observational Constraints with Specific Hubble Parameter, Symmetry, **15**, 1835 (2023).
[29] Sofuoglu, D., Baysal, H. and Tiwari, R.K. Observational constraints on the cubic parametrization of the deceleration parameter in f(R, T) gravity, Eur. Phys. J. Plus, **138**, 550 (2023).
[30] Dhruv Arora, Himanshu Chaudhary, Shibesh Kumar Jas Pacif, G. Mustafa, Diagnostic and Comparative Analysis of Dark Energy Models with q(z) Parametrizations, *https://arxiv.org/abs/2304.09749* (2023).
[31] A. Al Mamon, U. K. Sharma, M. Kumar et al., Cosmic consequences of Barrow holographic dark energy with Granda–Oliveros cut-off in fractal cosmology, Gen. Relativ. Gravit., **55**, 74 (2023).
[32] A. Chraya, Y. Muralichandran, G. Sethi, Variable Chaplygin gas: constraints from supernovae, GRB and gravitational wave merger events, Astrophys. Space Sci., **368**, 54 (2023).
[33] V. K. Bhardwaj, P. Garg, Exploring the cosmological model in $f(R, T^{\phi})$ gravity with observational constraints, New Astron., **105**, 102101 (2024).
[34] Shahnawaz A. Adil, Özgür Akarsu, Eleonora Di Valentino, Rafael C. Nunes, Emre Özülker, Anjan A. Sen, and Enrico Specogna, Omnipotent dark energy: A phenomenological answer to the Hubble tension, Phys. Rev. D, **109**, 023527 (2024).
[35] M. Khurana, H. Chaudhary, U. Debnath, A. Sardar, G. Mustafa, Exploring Late-Time Cosmic Acceleration with Eos Parameterizations in Horava-Lifshitz Gravity via Baryon Acoustic Oscillations, Fortschr. Phys., **72**, 2300238 (2024).
[36] G. A. Bogoslovsky, Viable Model of Locally Anisotropic Space-Time and the Finslerian Generalization of the Relativity Theory, Fortschr. Der Phys. Phys., **42**, 143–193 (1994).
[37] G. Gibbons, J. Gomis, C. Pope, General very special relativity is Finsler geometry, Phys. Rev. D, **76**, 081701 (2007).
[38] A. Kostelecky, Riemann-Finsler geometry and Lorentz-violating kinematics, Phys. Lett. B, **701**, 137–143 (2011).
[39] C. Pfeifer, Finsler spacetime geometry in Physics, Int. J. Geom. Methods Mod. Phys., **16**, 1941004 (2019).
[40] C. Lammerzahl, V. Perlick, Finsler geometry as a model for relativistic gravity, Int. J. Geom. Methods Mod. Phys., **15**, 1850166 (2018).
[41] S. Ikeda, E. N. Saridakis, P. C. Stavrinos, A. Triantafyllopoulos, Cosmology of Lorentz fiber-bundle induced scalar-tensor theories, Phys. Rev. D, **100**, 124035 (2019).

[42] A. Bouali, H. Chaudhary, R. Hama et al. Cosmological tests of the osculating Barthel–Kropina dark energy model, Eur. Phys. J. C, **83**, 121 (2023).
[43] R. Hama, T. Harko, S. V. Sabau, Dark energy and accelerating cosmological evolution from osculating Barthel–Kropina geometry, Eur. Phys. J. C, **82**, 385 (2022).
[44] R. Hama, T. Harko, S. V. Sabau et al., Cosmological evolution and dark energy in osculating Barthel–Randers geometry, Eur. Phys. J. C, **81**, 742 (2021).
[45] W. Feng, W. Yang, B. Jiang, Y. Wang, T. Han, Y. Wu, Theoretical analysis on the Barrow holographic dark energy in the Finsler–Randers cosmology, Int. J. Mod. Phys. D, **32**, 2350029 (2023).
[46] H. Lou, J. Li, W. Yang, W. Feng, W. Liu, Q. Zhang, N. Zhang, Y. Qi, Y. Wu, Theoretical analysis on the Rényi holographic dark energy in the Finsler–Randers cosmology, Int. J. Mod. Phys. D, **31**, 2250002 (2022).
[47] P. C. Stavrinos, A. P. Kouretsis, M. Stathakopoulos, Friedman-like Robertson–Walker model in generalized metric space-time with weak anisotropy, Gen. Relativ. Gravit., **40**, 1403–1425 (2008).
[48] Planck Collaboration, P. A. R. Ade et al., Planck 2015 results. XVI. Isotropy and statistics of the CMB, Astron. Astrophys., **594**, A16 (2016), arXiv:1506.07135 [astro-ph.CO].
[49] Planck Collaboration, N. Aghanim et al., Planck 2018 results. VII. Isotropy and Statistics of the CMB, Astron. Astrophys., **641**, A7 (2020), arXiv:1906.02552 [astro-ph.CO].
[50] C. L. Bennett et al., The Microwave Anisotropy Probe* Mission, Astrophys. J., **583**, 1 (2003).
[51] P. Bielewicz et al., Multipole vector anomalies in the first-year WMAP data: a cut-sky analysis, Astrophys. J., **635**, 750–760 (2005).
[52] C. L. Copi et al., On the large-angle anomalies of the microwave sky, Mon. Not. Roy. Astron. Soc., **367**, 79–102 (2005).
[53] G. F. R. Ellis, C. Hellaby, D. R. Matravers, Density waves in cosmology, Astrophys. J., **364**, 400–404 (1990).
[54] G. F. R., Ellis, R. Maartens, The Emergent Universe: inflationary cosmology with no singularity, Class. Quant. Grav., **21**, 223–232 (2004).
[55] D. R. Matravers, C. G. Tsagas, Cosmic magnetism, curvature, and the expansion dynamics, Phys. Rev. D, **62**, 103519 (2000).
[56] C. G. Tsagas, R. Maartens, Magnetized cosmological perturbations, Phys. Rev. D, **61**, 083519 (2000).
[57] P. K. Aluri et al., Is the Observable Universe Consistent with the Cosmological Principle?, Class. Quant. Grav., **40**(09), 094001 (2023).
[58] G. S. Asanov, Finsler Geometry, Relativity and Gauge Theories. Reidel, Dordrecht (1985).
[59] R. G. Beil, Electrodynamics from a metric, Int. J. Theor. Phys., **26**, 189–197 (1987).
[60] Z. Nekouee, S. K. Narasimhamurthy, H. M. Manjunatha, S. K. Srivastava, Finsler-Randers model for anisotropic constant-roll inflation, Eur. Phys. J. Plus, **137**, 1388 (2022).
[61] Z. Nekouee, S. K. Narasimhamurthy, Thermodynamic product formulae for Finslerian Kiselev black hole, Eur. Phys. J. C, **83** 723 (2023).
[62] P. C. Stavrinos, Congruences of fluids in a Finslerian anisotropic space-time, Int. J. Theor. Phys., **44**, 245–254 (2005).
[63] D. Bao, S. S. Chern, Z. Shen, An Introduction to Riemann-Finsler Geometry, Springer New York, (2000).
[64] I. W. Roxburgh, Finsler spaces with Riemannian geodesics, Gen. Relativ. Grav., **23**, 1071–1080 (1991).
[65] J. I. Horvarth, A possible geometrical interpretation of the isospace and of its transformations, Acta Phys. Chem. Szeged, **07**, 3–16 (1961).
[66] J. D. Bekenstein, Relation between physical and gravitational geometry, Phys. Rev. D, **48**, 3641 (1993).
[67] S. I. Vacaru, Locally anisotropic Kinetic processes and thermodynamics in curved spaces, Ann. Phys., **290**, 83–123 (2001).
[68] S. Vacaru, P. C. Stavrinos, E. Gaburov, D. Gonta, Clifford and Riemann- Finsler structures in geometric mechanics and gravity, Geometry Balkan Press, (2005).
[69] H. Rund, The Differential Geometry of Finlser Spaces, Springer Berlin, (1959).
[70] A. Triantafyllopoulos et al., Schwarzschild-like solutions in Finsler–Randers gravity, Eur. Phys. J. C, **80**, 1200 (2020).
[71] R. Raushan, R. Chaubey, Finsler–Randers cosmology in the framework of a particle creation mechanism: a dynamical systems perspective, Eur. Phys. J. Plus, **135**, 228 (2020).
[72] P. C. Stavrinos, A. P. Kouretsis, FRW-metric and Friedmann equations in a generalized cosmological model, J. Phys.: Conf. Ser., **68**, 012023 (2007).
[73] M. K. Roopa, S. K. Narasimhamurthy, On Finsler-cosmological models in Einstein and scalar-tensor theories, Palest. J. Math., **9**, 957–968 (2020).
[74] W. Zimdahl, D. Pavon, Letter: Statefinder Parameters for Interacting Dark Energy, Gen. Relativ. Gravit., **36**, 1483–1491 (2004).
[75] R. Raushan, A. K. Shukla, R. Chaubey, Aspects of non-interacting and interacting universe: a complete scenario, Can. J. Phys., **101**, 88–100 (2023).
[76] V. K. Oikonomou, I. Giannakoudi, A panorama of viable F(R) gravity dark energy models, Int. J. of Mod. Phys. D, **31**, 2250075 (2022).
[77] E. J. Copeland, M. Sami, S. Tsujikawa, Dynamics of dark energy, Int. J. Mod. Phys. D, **15**, 1753–1936 (2006).
[78] K. Bamba, S. Capozziello, Sh. Nojiri, Sergei D. Odintsov, Dark energy cosmology: the equivalent description via different theoretical models and cosmography tests, Astrophys. Space Sci., **342**, 155–228 (2012).
[79] J. V. Cunha, J. A. S. Lima, Transition Redshift: New Kinematic Constraints from Supernovae, Mon. Not. R. Astr. Soc., **390**, 210–217 (2008).
[80] E. Mortsell, C. Clarkson, Model independent constraints on the cosmological expansion rate, J. Cosm. Astropar. Phys., **2009**, 044 (2009).

[81] S. K. J. Pacif, R. Myrzakulov, Sh. Myrzakul, Reconstruction of cosmic history from a simple parametrization of H, Int. J. Geom. Meth. Mod. Phys., **14**, 1750111 (2017).

[82] S. Capozziello, R. Lazkoz, V. Salzano, Comprehensive cosmographic analysis by Markov Chain Method, Phys. Rev. D, **84**, 124061 (2011).

[83] M. Moresco, Raising the bar: New constraints on the Hubble parameter with cosmic chronometers at $z \sim 2$, Mon. Not. R. Astron. Soc., **450**(1), L16–L20 (2015).

[84] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, D. Wilkinson, et al., A 6% measurement of the Hubble parameter at $z = 0.45$: Direct evidence of the epoch of cosmic re-acceleration, J. Cosmol. Astropart. Phys., **2016**(05), 014 (2016).

[85] M. Moresco, L. Verde, L. Pozzetti, R. Jimenez, A. Cimatti, New constraints on cosmological parameters and neutrino properties using the expansion rate of the universe to $z = 1.75$, J. Cosmol. Astropart. Phys., **2012**(07), 053 (2012).

[86] M. Moresco, A. Cimatti, R. Jimenez, L. Pozzetti, G. Zamorani, M. Bolzonella, J. Dunlop, F. Lamareille, M. Mignoli, H. Pearce, et al., Improved constraints on the expansion rate of the universe up to $z = 1.1$ from the spectroscopic evolution of cosmic chronometers, J. Cosmol. Astropart. Phys., **2012**(08), 006 (2012).

[87] M. Smith, C. B. D'Andrea, M. Sullivan, A. Möller, R. C. Nichol, R. C. Thomas, A. G. Kim et al., First cosmology results using supernovae Ia from the dark energy survey: Survey overview, performance, and supernova spectroscopy, Astron. J., **160**(6), 267 (2020).

[88] C. Roberts, K. Horne, A. O. Hodson, A. D. Leggat, Tests of ΛCDM and conformal gravity using GRB and quasars as standard candles out to $z \sim 8$, arXiv preprint arXiv:1711.10369 (2017).

[89] M. Demianski, E. Piedipalumbo, D. Sawant, L. Amati, Cosmology with gamma-ray bursts-II. Cosmography challenges and cosmological scenarios for the accelerated universe, Astronomy & Astrophysics, **598**, A113 (2017).

[90] D. Benisty and D. Staicova, Testing late-time cosmic acceleration with uncorrelated baryon acoustic oscillation dataset, Astronomy & Astrophysics, **647**, A38 (2021).

[91] N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday et al., Planck 2018 results-VI. Cosmological parameters, Astronomy & Astrophysics, **641**, A6 (2020).

[92] W. Handley, M. Hobson, A. Lasenby, Polychord: Nested sampling for cosmology, Mon. Not. R. Astron. Soc., **450**(1), L61–L65 (2015).

[93] A. Lewis, Getdist: A Python package for analysing Monte Carlo samples, arXiv preprint arXiv:1910.13970 (2019).

[94] V. Sahni, et al. Statefinder-A new geometrical diagnostic of dark energy, JETP Lett., **77**, 201–206 (2003).

[95] U. Alam, et al., Exploring the expanding Universe and dark energy using the statefinder diagnostic, Mon. Not. R. Astron. Soc., **344**, 1057–1074 (2003).

[96] M. Sami et al., Cosmological dynamics of a nonminimally coupled scalar field system and its late time cosmic relevance, Phys. Rev. D, **86**, 103532 (2012).

[97] R. Myrzakulov, M. Shahalam, Statefinder hierarchy of bimetric and galileon models for concordance cosmology, J. Cosm. Astrop. Phys., **2013**(10), 047 (2013).

[98] V. Sahni, A. Shaifeloo, A. A. Starobinsky, Two new diagnostics of dark energy, Phys. Rev. D, **78**, 103502 (2008).

[99] C. Zunckel, C. Clarkson, Consistency Tests for the Cosmological Constant, Phys. Rev. Lett., **101**, 181301 (2008).

[100] M. Shahalam, S. Sami, A. Agarwal, *Om* diagnostic applied to scalar field models and slowing down of cosmic acceleration, Mon. Not. Roy. Astron. Soc., **448**, 2948–2959 (2015).

[101] A. Agarwal R. Myrzakulov, S. K. J. Pacif, M. Shahalam, Cosmic acceleration from coupling of baryonic and dark matter components: Analysis and diagnostics, Int. J. Mod. Phys. D, **28**, 1950083 (2019).

[102] S. I. Vacaru, On axiomatic formulation of gravity and matter field theories with MDRs and Finsler-Lagrange-Hamilton geometry on (co)tangent Lorentz bundles, arXiv:1801.06444 [physics.gen-ph] (2018).

[103] M. Kerscher, J. Weller, On model selection in cosmology, SciPost Physics Lecture Notes 009 (2019).

[104] N. J. Secrest, S. von Hausegger, M. Rameez, R. Mohayaee, S. Sarkar, A challenge to the standard cosmological model, Astrophys. J. Lett., **937**(2), L31 (2022).

[105] Y. Akrami, T. S. Koivisto, M. Sandstad, Cosmological constraints on ghost-free bigravity: Background dynamics and late-time acceleration, in: THE THIRTEENTH MARCEL GROSSMANN MEETING: On Recent Developments in Theoretical and Experimental General Relativity, Astrophysics and Relativistic Field Theories, World Scientific, 2015, pp. 1252–1254.

[106] H. Akaike, A new look at the statistical model identification, IEEE transactions on automatic control **19**(6), 716–723 (1974).

[107] G. Schwarz, Estimating the dimension of a model, Ann. Math. Stat., **6**, 461–464 (1978).

Indian Journal of Chemical Technology
Vol. 31, July 2024, pp. 631-641
DOI: 10.56042/ijct.v31i4. 9102

Curcumin and ZnO-NPs loaded xanthan gum and hydroxypropyl guar gum-based hydrogel film for food packaging

Ishita Parmar, Meenakshi Tanwar & Rajinder K. Gupta*

Department of Applied Chemistry, Delhi Technological University, Delhi-110042, India

*E-mail: rkg67ap@yahoo.com

Received 26 February 2024; accepted 22 May 2024

This study is focused on the synthesis of a hydrogel based film for food packaging applications. Xanthan gum (XG) and hydroxypropyl guar gum (HPGG) based hydrogel films have been synthesized using citric acid (CA) as crosslinker, loaded with zinc oxide (ZnO) nanoparticles (NPs) and curcumin (Cur). The structural characteristics of hydrogel films is analysed by FTIR and XRD. Thermal analysis is conducted using TGA and morphological examination is done using SEM. Rheological analysis is performed to determine the viscoelastic behaviour of the hydrogel films. Anti-microbial assay, anti-oxidant assay and SRB assay are conducted to analyse the anti-bacterial, scavenging and cytotoxicity activity of the hydrogel films. The film shows biodegradability under compost environment in 30 days. Fruit shelf-life enhancer capabilities are also assessed by using an edible coating of strawberry. The results so obtained showed that the synthesized XG-HPGG-CA-ZnO-Cur Film has potential application in edible food packaging.

Keywords: Curcumin, Citric acid, Hydroxypropyl guar gum, Xanthan gum, ZnO

Introduction

Food packaging materials ensure that the quality of food is maintained throughout its shelf life. Hydrogel is a beneficial approach for food packaging systems, especially in managing moisture, which can be a barrier to the longevity of food[1]. The primary purpose is to extend the life span of processed food products and maintain freshness, instead of being subjected to intricate and lengthy storage methods[2]. Due to their affordability and ability to withstand water exposure, plastic polymers have become a convenient and cost-effective choice for packaging. However, plastic films are not fully recyclable or biodegradable, leading environmentalists to look for materials derived from renewable resources[3].

Researchers have developed ecologically friendly packaging options, such as biopolymers, as a promising replacement for non-biodegradable plastics in food packaging[4]. Research is being conducted to enhance the performance of polymers by investigating superabsorbent hydrogel that can absorb water and other water-compatible fluids up to 100% of their dry weight[5]. Given the health risks associated with synthetic polymers and the surge in plastic waste, active packaging materials made from natural antimicrobial ingredients are gaining interest. Proteins, polysaccharides, and lipids are some diverse food-grade polymers being investigated for this purpose. Bio-based antimicrobial packaging can prevent undesired changes in food quality and ecological waste while carrying nutrients. These biodegradable polymers can be used as films or applied in the form of coatings to the food[6].

Bio-derived polymeric hydrogels offer an effective alternative for addressing the challenge of plastic packaging disposal in the context of food applications. Plasticused in food packaging, tends to be non-biodegradable, which poses an imminent threat to the environment. For the development of environmentally friendly packaging materials, a variety of renewable biopolymers, such as proteins, lipids, and their composites generated from plant and animal sources, have been studied as substitutes for their non-biodegradable petrochemical-based counterparts. In this study, a hydrogel-based film containing natural biopolymers is prepared to be used as a food packaging material. It also covers changes to the film's properties induced by biodegradation.

Xanthan gum (XG) is a high molecular weight acidic biopolymer obtained from (*Xanthomonas campestris)* during aerobic fermentation using corn or sugar cane derivatives. Its primary function is to thicken food and cosmetics due to its 2:2:1 molar ratio of d-glucosyl, d-mannosyl, and d-glucuronyl acid

residues along with varying amounts of pyruvyl groups[7]. XG exhibits hydrocolloid nature, making it highly functional in difficult settings such as high shear stress, high salt and acidic conditions. XG can also be used in small amounts to achieve zero-order release kinetics[8].

Another biopolymer with significant characteristics in food packaging is guar gum (GG), which is also a high molecular weight polysaccharide, obtained from *Cyamopsis tetragonolobus* embryos of the Leguminosae family. GG is composed of a polysaccharide chain consisting of mannose as the linear backbone and has a side chain made up of a [2:1] ratio of mannose to galactose[9]. Hydroxypropyl guar gum (HPGG), a derivative of GG, is created by reacting GG with propylene oxide. HPGG produces highly viscous and biodegradable solutions. HPGG is considered a better option in food packaging applications due to its greater thermal stability and solubility in comparison to GG[10].

Combining two biopolymers to synthesize hydrogels, increases their stability and usefulness[11]. Cross-linkers can enhance the mechanical properties of biopolymers, making them suitable for food packaging. Chemical crosslinking is the most effective method, particularly for biopolymers. Various crosslinking agents, such as Ca^{2+} ions, citric acid, glutaraldehyde and boric acid can be used[4]. Covalent bonds, hydrogen bonds, ionic interactions, hydrophobic and dipole-dipole interactions can all crosslink polymer chains, resulting in hydrogels with various shapes such as films, coatings and sheets[11]. XG and HPGG crosslinked with CA forms a film which is useful in the application of food preservation.

Further, zinc oxide (ZnO) nanoparticles (NPs) possess antioxidant activity and properties of food preservation. Along with ZnO NPs, curcumin (Cur) show antioxidant and antimicrobial properties. Both ZnO NPs and Cur can enhance the property of the XG-HPGG based film.

The aim of this research is to develop a novel hydrogel-based film of XG and HPGG cross-linked with CA and loaded with ZnO NPs and Cur for food packaging applications. Moreover, thermal analysis (TGA), structural characteristics (FTIR and XRD), morphology (SEM), rheological measurement (viscosity, amplitude and frequency sweep) mechanical strength, fruit preservation, anti-microbial activity, anti-oxidant activity and cytotoxicity assay were analysed for the hydrogel-based films.

Experimental Section

Materials

XG was purchased from local market, HPGG was procured from Hindustan pvt. Ltd. Bhiwadi, India, as a gifted sample, CA was purchased from CDH, New Delhi, Cur from SD fine chemical, Mumbai and ZnO NPs were purchased from sigma Aldrich, USA. Glycerol was purchased from Fisher chemicals, UK. Milli Q water was used for all the experimental work.

Synthesis of XG-HPGG-CA-ZnO-Cur hydrogel-based film

The food grade XG and HPGG were used for fabricating packaging hydrogel-based films. XG and HPGG being hydrophilic polymers easily dissolve in water. The ratio of [Gums: CA] was taken to be [1:1.5]. The required amount of XG and HPGG are dissolved in distilled water at a fixed interval of time. CA as a cross-linker was added to the homogenous solution of XG and HPGG. Glycerol was added as plasticizer to the homogenized solution. A fixed and same composition of [ZnO: Cur] was added to the hydrogel preparation with constant stirring for 3 h. Afterwards, sonication was done to remove entrapped air, then poured into Petri dish and left for drying in the oven at 45°C for 24 h. Subsequently, hydrogel-based films were made and then used for further characterization analysis. A control film i.e., XG-HPGG-CA was synthesized by same method without ZnO NPs and Cur.

Characterization of hydrogel-based film

The Perkin-Elmer model 2000 FT-IR was used to perform analysis of functional groups in XG, HPGG, XG-HPGG-CA film, and XG-HPGG-CA-ZnO-Cur film. The data was collected in the range of 4000-400 cm^{-1} by taking 32 scans at 4 cm^{-1} resolution. X-ray diffraction was conducted using a wide-angle X-ray diffractometer. Bruker D8 Advance was used with Cu-Kα radiations of λ = 1.5406 Å, a scanning range of 2θ (5°–80°), and a scan rate of 1°/min at 20 kV and a current of 10 mA. The SEM (JEOL JSM-6610LV) was utilized to examine the surface morphology of the samples at 500, 1000, and 2500X magnifications under an accelerating voltage of 20 kV. To determine the thermal stability of the samples, TGA (TGA 4000) was used. The samples were heated from 25°C to 800°C at a heating rate of 10°C/min in N_2 atmosphere. Using Anton Paar Modular Compact Rheometer 302 (MCR), the rheological properties of XG-HPGG-CA solution and XG-HPGG-CA-ZnO-Cur solutions were determined at room temperature. The tests were conducted using a 40 mm parallel plate.

The mechanical properties of XG-HPGG-CA-ZnO-Cur Film were analysed by a Universal Testing Machine (UTM), LLOYD LR 5K. Load range was 1-50 N with a 1000 mm extension range and 5 mm gauge length. The test and approach speed were 500 mm/min with no preload. A strip of 15 mm width and 150 mm length was cut from each film for mechanical testing. Under air conditions, three readings were taken for each sample and the average result was evaluated for mechanical properties of the film.

To evaluate the antimicrobial activity of the films, the agar disk diffusion method was used against *Escherichia coli* and *Staphylococcus aureus*, representing Gram-negative and Gram-positive bacterial strains respectively. Bacterial cultures were prepared in nutrient broth and inoculated onto agar plates. Disks were cut from each film and put on the plates. The fully formed zones (ffz) were measured using a digital calliper after incubating the plates for 14-16 h at 37°C. The films used in the experiment were 0.7 cm in diameter. In order to conduct a comparative study regarding the rate of food spoilage, fresh strawberries were taken. One strawberry was coated with synthesized hydrogel and the other was used in uncoated form and stored in glass petri plates. Samples were tested at regular intervals of 24 h for 1 week. All studies were conducted in natural lighting at a constant temperature of 25℃. The UV-visible spectrophotometer was used to evaluate the antioxidant activity of hydrogel solution by DPPH free-radical scavenging[12]. A modified method based on a previous study was used to determine the photochemical stability, in which varying amounts of hydrogel solution (3 mg, 5 mg, 7 mg, 20 mg, and 50 mg) of XG-HPGG-CA-ZnO-Cur solution were dissolved in 50 mL ethanol solution at 37°C[13]. Eq. (1) was used to determine the percentage of DPPH scavenging activity after mixing 1 mL of the solution with 3 mL of ethanol solution containing DPPH (1 mg/20 mL) and incubating it in the dark at room temperature for 30 min. The absorbance was measured at 517 nm.

$$\text{Scavenging activity of DPPH (\%)} = \left(1 - \frac{A_{sample}}{A_{control}}\right) x\ 100 \quad \ldots(1)$$

The value of $A_{control}$ indicates the absorbance of 1 mL ethanol and 3 mL ethanol solution of DPPH, while A_{sample} refers to the absorbance of 1 mL XG-HPGG-CA-ZnO-Cur solution and 3 mL DPPH solution, as stated in prior research[14].

The degradation tests were performed on XG-HPGG-CA-ZnO-Cur Film by the method of burring the film in standard soil (ISO 17556 standard soil) for 30 days at room temperature and moisture-controlled conditions. The film was buried 2 cm under the soil. The buried film was removed from the soil, gently brushed and washed with distilled water, dried and analysed for weight loss after every 5 days. The weight loss was observed through TGA. As a fragment of film was studied for weight loss and compared it with the freshly prepared film. The percentage weight loss was calculated as by Eq. (2),

$$\text{Percentage weight loss} = \left(\frac{w_i - w_f}{wi_i}\right) \text{x } 100 \quad \ldots(2)$$

Where w_i is the initial weight of the film and w_f is the weight after 5 days interval.

The cytotoxicity of the films using Hep-G2 cell line was determined using the SRB assay[15]. In the initial stage, cells were cultured in a 96-well plate with DMEM medium supplemented with 10% FBS and 1% antibiotic solution, at a density of 8000 cells/well, and incubated at 37°C with 5% CO_2. Following 24 h, the cells were exposed to varying concentrations of the dose (ranging from 0.05-1.25%), prepared in an incomplete medium. After an additional 24 h of incubation, each well was treated with 100 µL of 10% trichloro acetic acid (TCA) and incubated for 1 h. The plate was then washed with DM water and allowed to air dry at room temperature, followed by the addition of a final concentration of 0.04% SRB solution to each well, and incubated for 1 h. Once incubation was complete, the plate was rinsed with 1% (v/v) acetic acid to eliminate any unbound dye and left to air-dry at room temperature. To solubilize the protein-bound dye, a Tris base solution (pH=10.5) was introduced to each well and agitated for 10 min using an orbital shaker. Finally, the plate was examined in an ELISA plate reader (iMark, Biorad, USA) at 510 nm.

Results and Discussion

FTIR

FTIR was carried out on XG, HPGG, XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film to determine their functional groups in the range of 4000-400 cm^{-1} as shown in Fig. 1. XG showed characteristic peaks at 3325 cm^{-1} for O-H stretching and 1608 cm^{-1} for C=C stretching vibrations[16]. HPGG showed characteristic peaks at 2912 cm^{-1} for C-H stretching, 3342 cm^{-1} for N-H stretching and

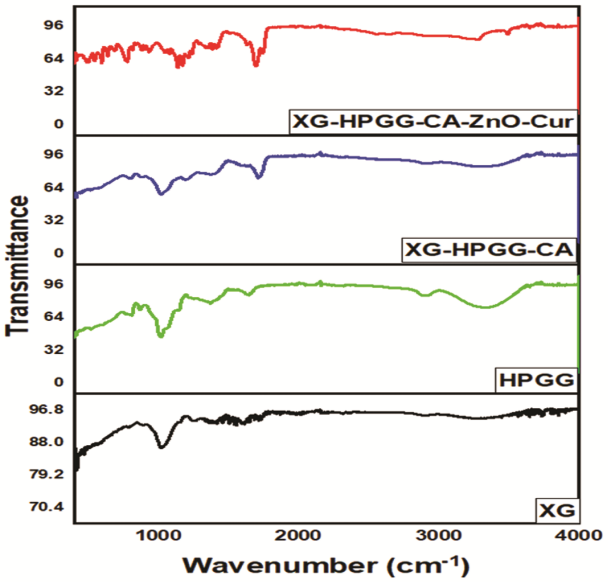


Fig. 1 — FTIR spectra of XG, HPGG, XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film

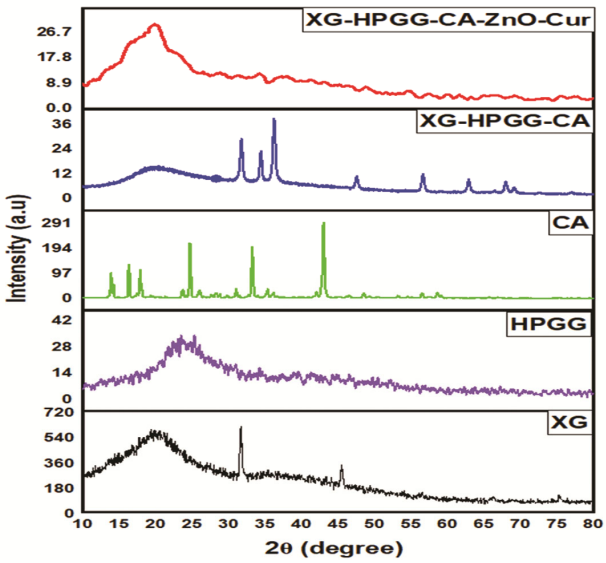


Fig. 2 — X-ray diffractogram of XG, HPGG, CA, XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film

1635 cm^{-1} for C=O stretching vibrations[17]. The FTIR spectra of XG-HPGG-CA showed a significant shift in peak from XG and HPGG at 3360 cm^{-1}, 2921 cm^{-1} and 1722 cm^{-1} due to crosslinking with CA. The FTIR spectra of XG-HPGG-CA-ZnO-Cur film showed a significant peak at 3273 cm^{-1}, corresponding to the -OH stretching vibrations, and 1697 cm^{-1}, due to carbonyl (C=O) stretching vibrations[18]. These shifts are majorly due to esterification with CA and incorporation of ZnO and Cur, which are essential in various applications, including food packaging[11].

XRD

The x-ray diffractogram of XG, HPGG, CA,XG-HPGG-CA film and XG-HPGG-CA-Cur-ZnO film are shown in Fig. 2. HPGG and XG exhibited broad peaks at 23.8° and 20.4°, respectively. CA showed peaks at 14º, 16º, 17º, 24.7º, 33º and 43º. XG-HPGG-CA film displayed a similar peak at 18.64° due to cross-linking of the two gums with CA. The shift of peaks was observed in XG-HPGG-CA-Cur-ZnO Film at 19.92º due to esterification. XRD patterns of XG, HPGG, XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film showed amorphous structure with low overall crystallinity. The crystalline regions of XG and HPGG were due to hydroxyl groups at18.86°and 19.96°, respectively. The stability of XG and HPGG was maintained by hydrogen bonds, which when broken, led to a decrease of crystallinity in XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film[19-22].

SEM

SEM images of XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film are shown in Fig. 3 at 500X and 1000X magnifications. SEM has been demonstrated to be quite useful for determining and verifying morphological traits[23,24]. The interior structure of the film was heterogeneous and inter-linked, which can result in significant swelling and permeability as well as facilitate cellular growth. Crosslinking of XG and HPGG with CA was responsible for homogeneous covering of XG-HPGG-CA film in Fig. 3c. On adding ZnO and Cur to the XG-HPGG-CA film, the morphology of the film changed from smooth surfaces to rough and hollow-follicular, as shown in Fig. 3(a,b). This shift in morphology after cross linking is due to the presence of ZnO and Cur constituents and the irregular pore network structure in XG-HPGG-CA-ZnO-Cur film is responsible for the heterogeneity in the polymeric hydrogel[25]. These structural characteristics provide additional evidence that cross-linked hydrogel networks were formed during the polymerization phase[26]. The surface morphology of polymers can provide valuable information regarding their specific pore size distribution or specific areas of the polymer matrix[14]. The structure of XG-HPGG-CA-ZnO-Cur film, as revealed by the surface morphology shows that this biopolymer is well suited to be used as a food packaging material[27].

TGA

TGA of XG, HPGG, CA, XG-HPGG-CA film, and XG-HPGG-CA-ZnO-Cur film is shown in Fig. 4. In

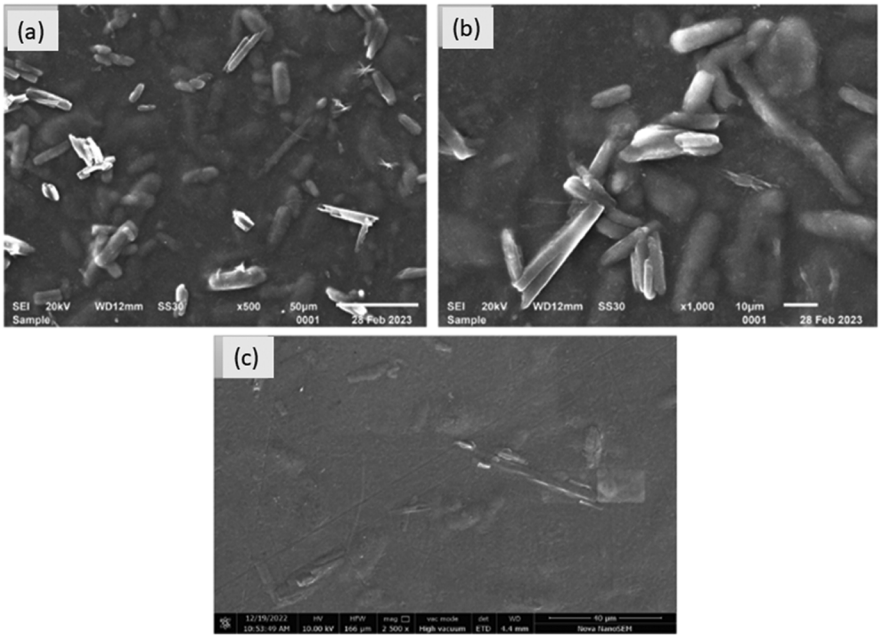


Fig. 3 — SEM micrographs at different magnifications of (a, b) XG-HPGG-CA-ZnO-Cur film at 500 and 1000X magnifications, respectively and (c) XG-HPGG-CA film at 2500X magnifications

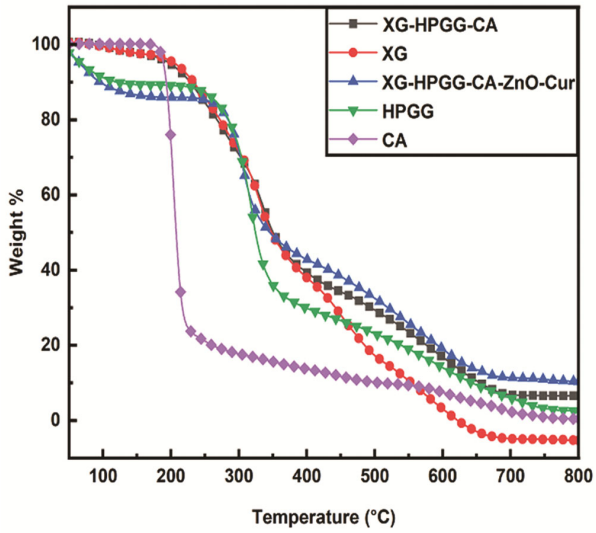


Fig. 4 — TGA of CA, HPGG, XG, XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film

each instance, the initial decomposition temperature (IDT) was employed to adjust for any loss of weight. For XG, an IDT of 13.66% was observed at (29°C to 222.27°C), while HPGG showed an IDT of 8.94% at (45.26°C to 183.92°C). IDT for CA was observed to be 83.772% at (163.68°C to 332.63°C). The crosslinked polymers, XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film, showed a weight loss of 3.239% and 2.268%, respectively, within almost the same temperature range. This demonstrates the added stability of XG-HPGG-CA-ZnO-Cur film was due to the presence of ZnO and Cur. T_{max} is the temperature at which the maximum rate of weight loss occurs[1], was observed in the range of (222.27°C to 418.32°C), (183.92°C to 392.92°C), (17.089°C to 388.9°C), and (151.2°C to 461.7°C) for XG, HPGG, XG-HPGG-CA film, and XG-HPGG-CA-ZnO-Cur film, respectively. These observations indicate that the thermal stability of gums, such as XG and HPGG, decreases during the stages of decomposition, but the polymerization and formation of films of XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film, showed an increased degree of stability due to the less percentage of weight loss compared to XG and HPGG[28].

Viscosity

Rheology is the scientific study of the deformation and flow of matter, which elucidates the relationship between force, time and deformation. It is the experimental method employed to assess the rheological characteristics of materials[29]. Shear rheological testing was conducted by varying the shear rate from 10 s^{-1} to 100 s^{-1}. Throughout the shear rate range, it was observed that XG-HPGG-CA-ZnO-Cur solution exhibited a stronger shear thinning behaviour than XG-HPGG-CA solution as shown in Fig. 5a. The significant shear-thinning characteristics were attributed to the incorporation of ZnO and Cur in

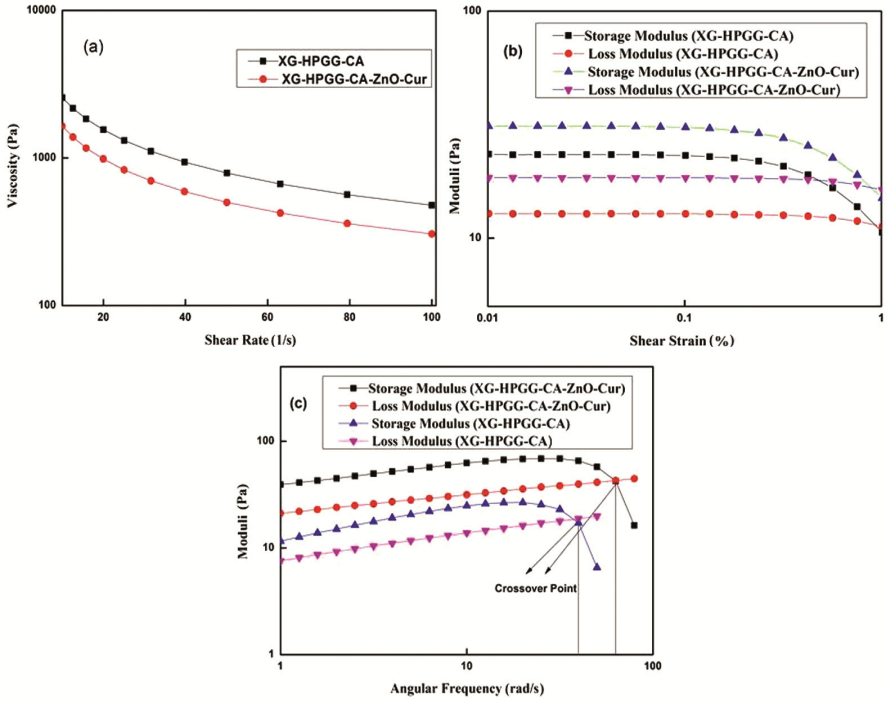


Fig. 5 — (a) Shear viscosity profile (b) amplitude sweep and (c) frequency sweep of XG-HPGG-CA solution and XG-HPGG-CA-ZnO-Cur solution

the XG-HPGG-CA-ZnO-Cur solution composition. In both cases, a considerable reduction in viscosity was observed due to the shear-induced alignment of XG-HPGG-CA solution and XG-HPGG-CA-ZnO-Cur solution in the flow direction[30].

Amplitude sweep

The amplitude sweep was conducted to determine the upper and lower limits of the linear viscoelastic (LVE) range and to assess the hydrogel's structural and mechanical stability[31]. Fig. 5b illustrates the plotted results of the storage modulus G' and loss modulus G" against shear strain percentage. The LVE range was characterized by a constant G' at low deformation, indicating that the sample structure remained intact[32]. At lower strain, both XG-HPGG-CA-ZnO-Cur solution and XG-HPGG-CA solution behaved as viscoelastic solids, while at higher strain, they behaved as viscoelastic fluids due to the disturbance of the structure, indicating the end of the LVE-region. The sample's rigidity at rest is characterized by the plateau value G′ within the LVE region. The LVE range for XG-HPGG-CA-ZnO-Cur solution was found to be G′=29.8 Pa, and for XG-HPGG-CA solution, G′=22.9 Pa, due to the addition of Cur and ZnO. When the sample reaches its most solid state, the LVR typically ends. The shear modulus at this point was observed to recover around 80% of its initial value[33]. The solution with ZnO and Cur showed a lower critical value (γc) of 18.6% compared to the solution with XG-HPGG-CA (γc) of 13%.

Frequency sweep

The frequency sweep test is used to assess the viscoelastic properties of a material by comparing the values of G' and G" over a range of frequencies[34]. Ajovalasit *et* al. conducted a study to investigate the effects on G' and G" within a frequency range by adding additives and concluded that all hydrogels showed the viscoelastic nature and found that both G' and G" showed positive slopes with the loss modulus increasing more rapidly compared to the storage modulus[35]. Throughout the test, the shear strain percentage was kept constant in the LVE region and frequency varied from 1 to 100 rad/s. The results in Fig. 5c show that for both XG-HPGG-CA solution and XG-HPGG-CA-ZnO-Cur solution solutions, G" dominates at high frequency while G' dominates at low frequency, indicating a viscoelastic solid behaviour and the Moduli of XG-HPGG-CA-ZnO-Cur solution is higher than XG-HPGG-CA hydrogel solution at all frequencies which proves extra stability due to ZnO and Cur.

Table 1 — Mechanical properties

Film	Maximum force (N)	Tensile strength (MPa)	Elongation at break (%)
XG-HPGG-CA film	49.5 ± 0.54	28.4 ± 0.10	8.9 ± 0.24
XG-HPGG-CA-ZnO-Cur film	54.6 ± 0.42	29.7 ± 0.15	7.88± 0.54

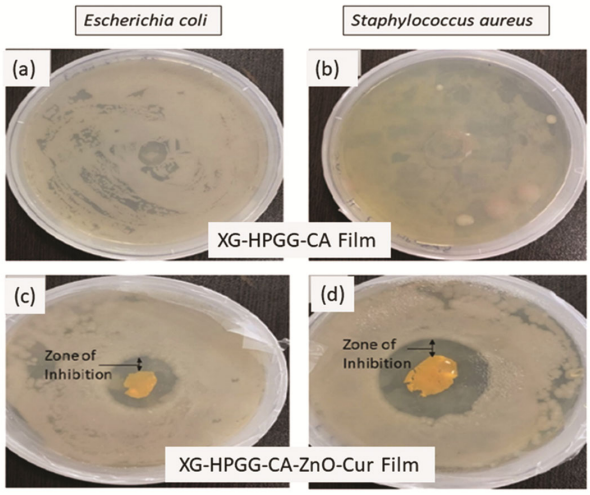


Fig. 6 — Disk diffusion test against *Escherichia coli* (a and c) and *Staphylococcus aureus* (b and d) for XG-HPGG-CA Film (a ,b) and XG-HPGG-CA-ZnO-Cur Film (c, d).

Mechanical Properties

The tensile properties of XG-HPGG-CA and XG-HPGG-CA-ZnO-Cur films were studied. The results for thickness (mm), width (mm), maximum load (N), elongation at break (%) and tensile strength (MPa) of the film are shown in the Table 1.

The XG-HPGG-CA-ZnO-Cur film showed an elongation at break % of 7.88 ± 0.54 which indicated its ability to withstand stretching and deformation before breaking. It has a high tensile strength of 29.7 ± 0.15 MPa which was indicative of its ability to withstand tension and resist breaking under force. These properties have made this film suitable for biomedical applications such as food packaging material.

***In Vitro* Test**

Disk Diffusion test

Active food packaging technology, involving incorporation of antimicrobial agents in the packaging to prevent microbial growth and prolong the shelf life of food, is becoming popular[36]. Fig. 6 shows the results of testing the antimicrobial activity of XG-HPGG-CA film and XG-HPGG-CA-ZnO-Cur film composites against two bacterial pathogen strains, *Escherichia coli* and *Staphylococcus aureus* using the disk diffusion method. In Fig. 6(a,b), XG-HPGG-CA film did not exhibit any antibacterial activity against either *E. coli* or *S. aureus* and it could not penetrate the bacterial cell. However, XG-HPGG-CA-ZnO-Cur film, (Fig. 6 (c, d) demonstrated a clear zone of inhibition of 87 mm and 102 mm, respectively, against both *E. coli*. and *S. aureus*. The results imply that XG-HPGG-CA-ZnO-Cur film exhibits significant antibacterial properties against both Gram-negative and Gram-positive bacteria, indicating that it has a broad-spectrum antibacterial activity.

The presence of an inhibition zone in the XG-HPGG-CA-ZnO-Cur film nanocomposite suggests that the biocidal action of ZnO NPs involves disrupting the membrane by generating surface oxygen species rapidly, ultimately leading to the death of pathogens[37]. This demonstrates the potential of XG-HPGG-CA-ZnO-Cur film composites as an effective active food packaging technology for inhibiting microbial growth and extending the shelf life of food products.

Food sampling test

After conducting antimicrobial tests using the disk diffusion method, the films were evaluated as a potential food packaging material for a food sample coated with hydrogel and a comparison was made with non-coated food material. To create a protective layer over the surface, the strawberries were directly immersed in the XG-HPGG-CA-ZnO-Cur solution and left to air dry[38]. Although strawberries have high antioxidant properties, their susceptibility to mechanical damage can lead to infectious symptoms, and their postharvest life is limited to about five days at 4°C[39]. The study aimed to enhance the shelf life of strawberries by coating them with an effective solution of prepared hydrogel.

Fig. 7 depicts one strawberry coated with the XG-HPGG-CA-ZnO-Cur solution while the other remains uncoated. The XG-HPGG-CA-ZnO-Cur solution coated strawberry showed slower deterioration compared to the uncoated sample. Day-1, showed no microbial growth on either of the fruit samples. Day 3, showed microbial growth on the non-coated fruit sample, and the fruit was shown to be deteriorated by Day 7. On the other hand, no microbial growth is observed on the strawberries coated with XG-HPGG-CA-ZnO-Cur solution coating till day 7. All these results confirm XG-HPGG-ZnO-Cur solution as a potential food packaging material with antimicrobial characteristics and Cur and ZnO as effective ingredients in active packaging films. Fruits and

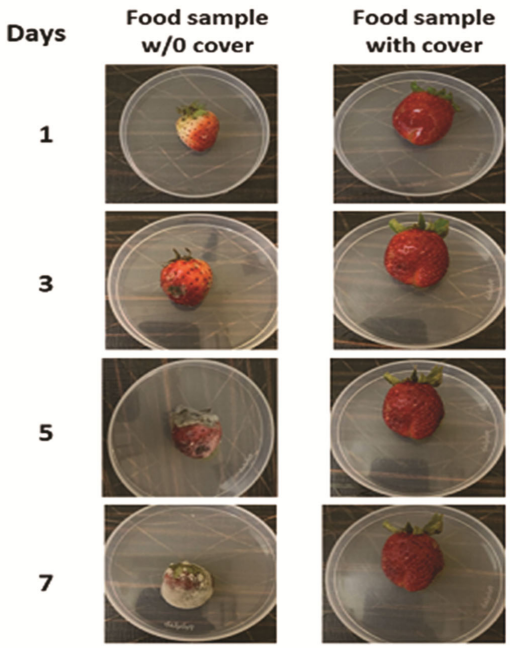


Fig. 7 — Food sample testing using strawberry without coating and with coating of XG-HPGG-CA-ZnO-Cur solution

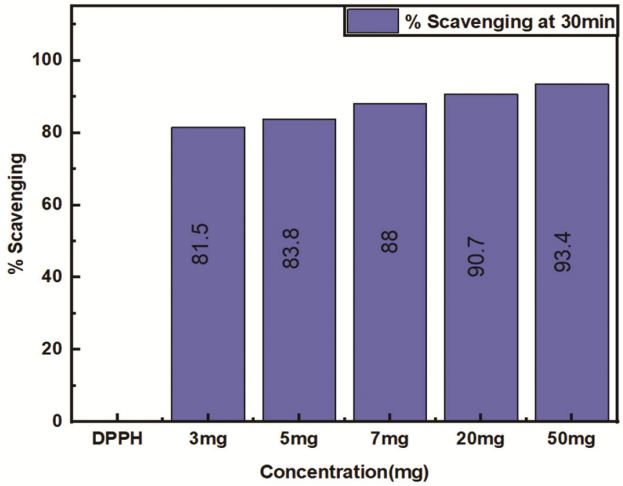


Fig. 8 — Scavenging of $DPPH^+$ radical by XG-HPGG-CA-ZnO-Cur film

vegetables with a high respiration rate can also be packaged using biopolymer sheets with a water permeability function[40].

Antioxidant activity

Although the conjugated diene moiety found in Cur accounts for the majority of its activities, the nearby phenolic groups have also been implicated in its antioxidant properties in Fig. 8. The results of the $DPPH^+$ Assay demonstrated that Cur facilitated hydrogen atom transfer, leading to conversion of a

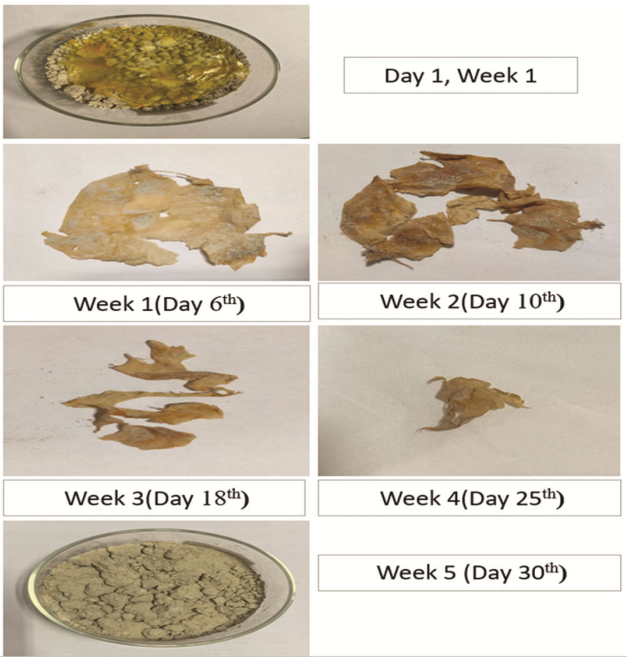


Figure. 9 — Biodegradation of XG-HPGG-CA-ZnO-Cur film

Table 2 — Biodegradability assay: XG-HPGG-CA-ZnO-Cur film weight loss

Day	Initial weight (g)	Final weight (g)	Percentage weight loss
5	0.472	0.337	28.6%
10	0.315	0.194	38.4%
15	0.194	0.093	52%
20	0.093	0.039	58%
25	0.039	0.085	78.2%
30	0.085	0.000	100%

stable radical called 2,2-diphenyl-1-picryl hydrazyl into 2,2-dipheny-l-picryl hydrazine (diamagnetic, non-radical)[41]. The assay showed a trend of increasing antioxidant activity. The combination of ZnO nanoparticles and Cur exhibited significant antioxidant activity against $DPPH^+$ free radicals[42]. The antioxidant activity was evaluated by incubating the solutions of varying concentrations (3 mg, 5 mg, 7 mg, 20 mg, 50 mg) with $DPPH^+$ solution for 30 min, resulting in an increase in antioxidant activity of 81.5%, 83.8%, 88%, 90.7% and 93.4%, respectively. The antioxidant activity of food packaging film can positively impact the packaging process by regulating the production of reactive oxygen species[43].

Biodegradability test using soil burial method

The results of the biodegradability test using soil burial method have been shown in the Fig. 9 and degradation was calculated at the interval of every 5 days that was shown in Table 2. It has been

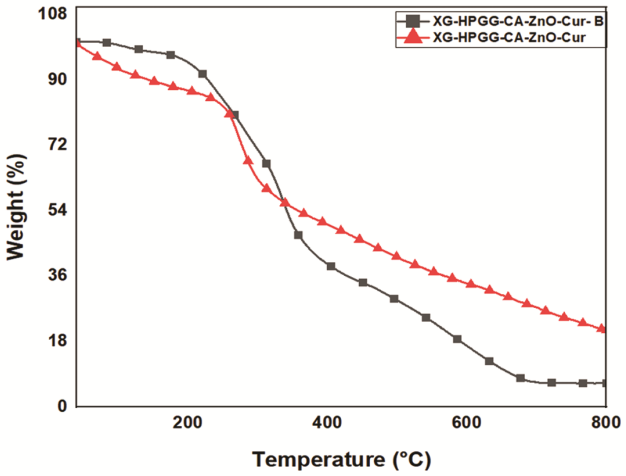


Fig. 10 — TGA graph of initial XG-HPGG-CA-ZnO-Cur film and the same film upon biodegradation (XG-HPGG-CA-ZnO-Cur –B)

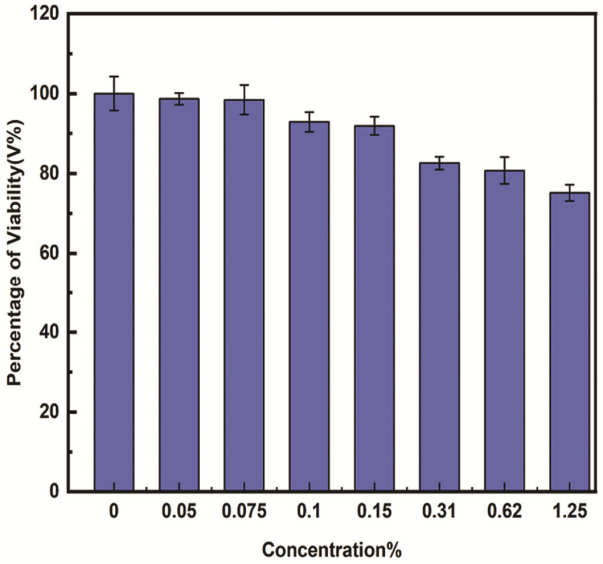


Fig. 11 — The viability percentage of Hep-G2 cells, for XG-HPGG-CA-ZnO-Cur film

observed that longer the burial time, higher will be the degradation of the film, which proved the environmentally friendly properties of the film[44]. The film gradually started to break down into small fragments around day-6 of first week after burial. Deterioration of film was seen on the surface as the film became fragile further in week 2. Cracks were observed on the surface of the sample all along the coming weeks of week 3 and week 4 as the film broke down in further smaller fragments. Soil degradation influenced macrostructural changes. Most of the film was degraded by the end of week 4 and only a small section of the film could be seen. The film was seen to be completely degraded around day 30th in week 5.

The film was weighed to calculate the weight loss over the period of degradation of 5 days. On around 5th day in week 1, the film was dusted and washed with distilled water to perform thermal degradation analysis. The results were then compared with an initial bioactive film as shown in Fig. 10. Shifts in peaks of both the films and amount of residual content was observed which further led to the observation that components of the biodegraded film broke down due to natural decomposition of the film upon microbial action in the soil. The biodegraded film showed wt. loss at a much faster rate than the XG-HPGG-CA-ZnO-Cur Film. Hence, it could be concluded that XG-HPGG-CA-ZnO-Cur Film is biodegradable film, which is an important aspect in packaging applications.

Cytotoxicity evaluation by SRB Essay

The cytocompatibility of XG-HPGG-CA-ZnO-Cur films was assessed using human hepatic cell lines (Hep-G2) as model cell lines. The results of the cytotoxicity test for different concentrations are presented in Fig. 11. The experiment with Hep-G2 cell culture demonstrated that the cells were able to attach to the surfaces of the films, proliferate, and disseminate[15]. After the incubation period, the cells on the XG-HPGG-CA-ZnO-Cur films exhibited normal morphology. The viability values of all film samples at different concentrations as shown in Fig. 11 were above 80%, indicating that the films were biocompatible with Hep-G2 cells.

Conclusion

This study demonstrated the successful fabrication of XG-HPGG-CA-ZnO-Cur films with antimicrobial properties and a strong zone of inhibition, making them suitable for use as edible food packaging material. The addition of glycerol also improved their thermal and mechanical stability compared to XG-HPGG-CA-ZnO-Cur. The characterization of the samples using various analytical techniques, such as FTIR, XRD, SEM, TGA, rheology, *in-vitro* tests revealed that XG-HPGG-CA-ZnO-Cur films promoted cell growth and differentiation, indicating their potential as a food packaging material. The films have antimicrobial, antioxidant properties and show biodegradable properties. These findings suggest that the use of harmful plastics in food packaging could potentially be reduced by utilizing these edible and affordable options. Overall, XG-HPGG-CA-ZnO-Cur films show promise as a sustainable and biocompatible alternative for food packaging applications.

Acknowledgements

The authors would like to thank Dr. Swarmita Dixit (Aakaar Biotechnologies Pvt. Ltd., Jankipuram Extension, Lucknow, India) for carrying out SRB test for the synthesized XG-HPGG-CA-ZnO-Cur Film. The authors are also thankful to UFlex Pvt. Ltd. Noida, India, for mechanical strength testing. Furthermore, the authors acknowledge the Vice Chancellor of DTU, for supporting the project with research facilities and financial aid.

References

1 Kang M, Oderinde O, Liu S, Huang Q, Ma W, Yao F & Fu G, Characterization of Xanthan gum-based hydrogel with Fe3+ ions coordination and its reversible sol-gel conversion, *Carbohydr Polym*, 203 (2019) 139.

2 Mohsin A, Zaman W Q, Guo M, Ahmed W, Khan I M, Niazi S, Rehman A, Hang H & Zhuang Y, Xanthan-curdlan nexus for synthesizing edible food packaging films, *Int J Biol Macromol*, 162 (2020) 43.

3 Farris S, Schaich K M, Liu L S, Piergiovanni L & Yam K L, Development of polyion-complex hydrogels as an alternative approach for the production of bio-based polymers for food packaging applications: A review, *Trends Food Sci Technol*, 20 (2009) 316.

4 Garavand F, Rouhi M, Razavi S H, Cacciotti I & Mohammadi R, Improving the integrity of natural biopolymer films used in food packaging by crosslinking approach: A review, *Int J Biol Macromol*, 104 (2017) 687.

5 Batista R A, Espitia P J P, Quintans J S S, Freitas M M, Cerqueira M A, Teixeira J A & Cardoso J C, Hydrogel as an alternative structure for food packaging systems, *Carbohydr Polym*, 205 (2019) 106.

6 Kaur M & Santhiya D, UV-shielding antimicrobial zein films blended with essential oils for active food packaging, *J Appl Polym Sci*, 138 (2021) 1.

7 Bueno V B, Bentini R, Catalani L H & Petri D F S, Synthesis and swelling behavior of xanthan-based hydrogels, *Carbohydr Polym*, 92 (2013) 1091.

8 Malik N S, Ahmad M, Minhas M U, Tulain R, Barkat K, Khalid I & Khalid Q, Chitosan/xanthan gum based hydrogels as potential carrier for an antiviral drug: Fabrication, characterization, and safety evaluation, *Front Chem*, 8 (2020) 50.

9 Zhu J, Guan S, Hu Q, Gao G, Xu K & Wang P, Tough and pH-sensitive hydroxypropyl guar gum/polyacrylamide hybrid double-network hydrogel, *Chem Eng J*, 306 (2016) 953.

10 Parameswaran-Thankam A, Al-Anbaky Q, Al-karakooly Z, Rangu-Magar A B, Chhetri B P, Ali N & Ghosh A, Fabrication and characterization of hydroxypropyl guar-poly (vinyl alcohol)-nano hydroxyapatite composite hydrogels for bone tissue engineering, *J Biomater Sci Polym Edn*, 29 (2018) 2083.

11 Simões B M, Cagnin C, Yamashita F, Olivato J B, Garcia P S, de Oliveira S M & Grossmann M V E, Citric acid as crosslinking agent in starch/xanthan gum hydrogels produced by extrusion and thermopressing, *LWT-Food Sci Technol*, 125 (2020) 108950.

12 Ma Q, Ren Y & Wang L, Investigation of antioxidant activity and release kinetics of curcumin from tara gum/ polyvinyl alcohol active film, *Food Hydrocoll*, 70 (2017) 286.

13 Sun C, Xu C, Mao L, Wang D, Yang J & Gao Y, Preparation, characterization and stability of curcumin-loaded zein-shellac composite colloidal particles, *Food Chem*, 228 (2017) 656.

14 Wang H, Gong X, Guo X, Liu C, Fan Y Y, Zhang J, Niu B & Li W, Characterization, release, and antioxidant activity of curcumin-loaded sodium alginate/ZnO hydrogel beads, *Int J Biol Macromol*, 121 (2019) 1118.

15 Kianpour S, Ebrahiminezhad A, Mohkam M, Tamaddon A M, Dehshahri A, Heidari R & Ghasemi Y, Physicochemical and biological characteristics of the nanostructured polysaccharide-iron hydrogel produced by microorganism Klebsiella oxytoca, *J Basic Microbiol*, 57 (2017) 132.

16 Zhang Q, Hu X M, Wu M Y, Wang M M, Zhao Y Y & Li T T, Synthesis and performance characterization of poly(vinyl alcohol)-xanthan gum composite hydrogel, *React Funct Polym*, 136 (2019) 34.

17 Lu X, Li Y, Feng W, Guan S & Guo P, Self-healing hydroxypropyl guar gum/poly (acrylamide-co-3-acrylamidophenyl boronic acid) composite hydrogels with yield phenomenon based on dynamic PBA ester bonds and H-bond, *Colloids Surfaces A Physicochem Eng Asp*, 561 (2019) 325.

18 Nawaz A, Farid A, Safdar M, Latif M S, Ghazanfar S, Akhtar N, Al-Jaouni S K, Selim S & Khan M W, Formulation development and Ex-vivo permeability of curcumin hydrogels under the influence of natural chemical enhancers, *Gels*, 8 (2022) 1.

19 Tanwar R, Kumar L, Kumar P & Gaikwad K K, Characterization of guar gum based edible film containing microwave-assisted mint leaves extract and citric acid for ber (Ziziphus Mauritiana) fruit packaging, *SSRN Electron J*, (2021) 3978979.

20 Yanhan Q, Shucai L, Zhaofeng L, Jian Z & Haiyan L, Hydration effect of sodium silicate on cement slurry doped with xanthan, *Constr Build Mater*, 223 (2019) 976.

21 Katherine R F, Muthukumaran C, Sharmila G, Kumar N M, Tamilarasan K & Jaiganesh R, Xanthan gum production using jackfruit-seed-powder-based medium: Optimization and characterization, *3 Biotechnol*, 7 (2017) 248.

22 Chen X, Li P, Kang Y, Zeng X, Xie Y, Zhang Y, Wang Y & Xie T, Preparation of temperature-sensitive Xanthan/NIPA hydrogel using citric acid as crosslinking agent for bisphenol A adsorption, *Carbohydr Polym*, 206 (2019) 94.

23 Horn M M, Martins V C A & de Guzzi P A M, Influence of collagen addition on the thermal and morphological properties of chitosan/xanthan hydrogels, *Int J Biol Macromol*, 80 (2015) 225.

24 Williams P D, Oztop M H, Mccarthy M J, Mccarthy K L & Lo Y M, Characterization of water distribution in xanthan-curdlan hydrogel complex using magnetic resonance imaging, nuclear magnetic resonance relaxometry, rheology, and scanning electron microscopy, *J Food Sci*, 76 (2011) 472.

25 Singh B & Singh B, Graft copolymerization of polyvinylpyrollidone onto Azadirachta indica gum polysaccharide in the presence of crosslinker to develop hydrogels for drug delivery applications, *Int J Biol Macromol*, 159 (2020) 264.

26 Singh B, Sharma S & Dhiman A, Acacia gum polysaccharide based hydrogel wound dressings: Synthesis, characterization, drug delivery and biomedical properties, *Carbohydr Polym*, 165 (2017) 294.

27 Manna P J, Mitra T, Pramanik N, Kavitha V, Gnanamani A & Kundu P P, Potential use of curcumin loaded carboxymethylated guar gum grafted gelatin film for biomedical applications, *Int J Biol Macromol*, 75 (2015) 437.

28 Singh B & Sharma N, Mechanistic implication for cross-linking in sterculia-based hydrogels and their use in GIT drug delivery, *Biomacromolecules*, 10 (2009) 2515.

29 Liepsch D, A basic introduction to rheology shear flow, *J Biomech*, 35 (2016) 415.

30 Wang C S, Virgilio N, Carreau P J & Heuzey M C, Understanding the effect of conformational rigidity on rheological behavior and formation of polysaccharide-based hybrid hydrogels, *Biomacromolecules*, 22 (2021) 4016.

31 Merlusca I P, Ibanescu C, Tuchilus C, Danu M, Atanase L I & Popa I M, Characterization of neomycin-loaded xanthan-chitosan hydrogels for topical applications, *Cellul Chem Technol*, 53 (2019) 709.

32 Yadav R & Purwar R, Influence of metal oxide nanoparticles on morphological, structural, rheological and conductive properties of mulberry silk fibroin nanocomposite solutions, *Polym Test*, 93 (2021) 106916.

33 Herrada-Manchón H, Rodríguez-González D, Fernández M A, Kucko N W, De G F B & Aguilar E, Effect on rheological properties and 3D printability of biphasic calcium phosphate microporous particles in hydrocolloid-based hydrogels, *Gels*, 8 (2022) 28.

34 Stojkov G, Niyazov Z, Picchioni F & Bose R K, Relationship between structure and rheology of hydrogels for various applications, *Gels*, 7 (2021) 255.

35 Ajovalasit A, Sabatino M A, Todaro S, Alessi S, Giacomazza D, Picone P, Carlo M D & Dispenza C, Xyloglucan-based hydrogel films for wound dressing: Structure-property relationships, *Carbohydr Polym*, 179 (2018) 262.

36 Raschip I E, Fifere N, Varganici C D & Dinu M V, Development of antioxidant and antimicrobial xanthan-based cryogels with tuned porous morphology and controlled swelling features, *Int J Biol Macromol*, 156 (2020) 608.

37 Rukmanikrishnan B, Ismail F R M, Manoharan R K, Kim S S & Lee J, Blends of gellan gum/xanthan gum/zinc oxide based nanocomposites for packaging application: Rheological and antimicrobial properties, *Int J Biol Macromol*, 148 (2020) 1182.

38 Valdés A, Ramos M, Beltrán A, Jiménez A & Garrigós M C, State of the art of antimicrobial edible coatings for food packaging applications, *Coatings*, 7 (2017) 1.

39 Liu Y, Wang S, Lan W & Qin W, Fabrication and testing of PVA/Chitosan bilayer films for strawberry packaging, *Coatings*, 7 (2017) 1.

40 Wang L F & Rhim J W, Preparation and application of agar/alginate/collagen ternary blend functional food packaging films, *Int J Biol Macromol*, 80 (2015) 460.

41 Bhoopathy S, Inbakandan D, Rajendran T, Chandrasekaran K, Kasilingam R & Gopal D, Curcumin loaded chitosan nanoparticles fortify shrimp feed pellets with enhanced antioxidant activity, *Mater Sci Eng C*, 120 (2021) 111737.

42 Suresh D, Shobharani R M, Nethravathi P C, Kumar P M A, Nagabhushana H & Sharma S C, Artocarpus gomezianus aided green synthesis of ZnO nanoparticles: Luminescence, photocatalytic and antioxidant properties, *Spectrochim Acta-Part A Mol Biomol Spectrosc*, 141 (2015) 128.

43 Singh B & Rajneesh, Gamma radiation synthesis and characterization of gentamicin loaded polysaccharide gum based hydrogel wound dressings, *J Drug Deliv Sci Technol*, 47 (2018) 200.

44 Nissa R C, Fikriyyah A K, Abdullah A H D & Pudjiraharti S, Preliminary study of biodegradability of starch-based bioplastics using ASTM G21-70, dip-hanging, and soil burial test methods, *IOP Conf Ser Earth Environ Sci*, 277 (2019) 012007.

Indian J Phys
https://doi.org/10.1007/s12648-024-03343-9

ORIGINAL PAPER

Design and theoretical study of rectangular photonic crystal fiber based sensor for chemical sensing in terahertz regime

D Garg, J Singh and A Kumar*

Advanced Photonics Simulation Research Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi, Delhi 110042, India

Received: 18 April 2024 / Accepted: 05 July 2024

Abstract: This paper presents the design and theoretical analysis of a photonic crystal fiber (PCF) based chemical sensor model. To evaluate the efficiency of this model, various optical parameters are analysed using finite element method based COMSOL Multiphysics Software. Different analytes namely methanol (1.317), water (1.330), ethanol (1.354) and benzene (1.366) have been considered for the sensing purpose in this study. The core region is infiltrated with different analytes separately. The model is simulated in THz regime (0.5–1.5 THz) to evaluate optical properties. The proposed structure design exhibits a high relative sensitivity of 96.85%, 97.32%, 97.99% and 98.33% for methanol, water, ethanol, and benzene, respectively at an operating frequency of 1.3 THz. The proposed model demonstrates exceptionally low confinement loss values which are 2.22×10^{-12} dB/m for methanol, 1.16×10^{-11} dB/m for water, 1.34×10^{-11} dB/m for ethanol and is 1.30×10^{-12} dB/m for benzene. Additionally, the effective material loss for the designed PCF also comes out to be very low for all the analytes, 0.0044 cm^{-1} for methanol, 0.0040 cm^{-1} water, 0.0034 cm^{-1} for ethanol and 0.0032 cm^{-1}, for benzene. Furthermore, the PCF shows large effective mode area and numerical aperture (NA) within the mentioned range, at 1.3 THz. The NA values obtained at 1.3 THz are 0.32 for methanol, 0.40 for water, 0.32 for ethanol, and 0.32 for benzene. The obtained Effective mode Area (EMA) values are 1.46×10^5 μm^2 for methanol, 1.45×10^5 μm^2 for water, 1.44×10^5 μm^2 for ethanol and 1.43×10^5 μm^2, for benzene. Subsequently, the optimal profile provides birefringence values of 0.0009 for methanol, 0.0010 for water, and 0.0011 for both ethanol and benzene. The practical implementation of the proposed PCF structure is possible using subsisting modern fabrication techniques. Consequently, the proposed PCF design should be helpful in industrial areas for chemical research, food and biomedical sensing.

Keywords: Photonic crystal fiber; Terahertz; Relative sensitivity; Confinement loss; Chemical sensing

1. Introduction

Analysing and exploring unidentified substances is an intriguing subject in the field of science, essential for the well-being of human health. For this purpose, researchers are working diligently to develop and design different kinds of sensors which can aid in detecting unknown chemicals. Substances like alcohol and water are the most common chemicals that are used as primary components in many compound solutions. So, an effective method should be accessible for chemical sensing. Sensors based on PCFs are expected to be beneficial in industrial sectors, particularly for applications in chemical research, as well as in the fields of food and biomedical sensing [1]. In this paper, we have explored the application of chemical sensing in the terahertz regime. In the past decade, photonic crystal fibers (PCFs) operating in the terahertz region have gained tremendous popularity. Terahertz frequency belongs to the sect of electromagnetic radiation that lies in the range of microwave and infrared frequency. Terahertz radiation frequency region ranges from 0.1 to 10 THz [2]. A major setback that hindered the comprehensive study of the THz band earlier was the transmission loss due to the atmospheric moisture during propagation [3]. To address this problem in THz communications, THz optical photonic crystal fibers (PCFs) have been utilized as waveguides across a range of applications, encompassing THz antennas [4] and polarization-maintaining THz systems [5]. PCFs operating in THz regime showcase numerous promising traits and advantages; THz sensors are bigger than the sensors operating in the mid-infrared region, in terms of

*Corresponding author, E-mail: ajeetdph@dtu.ac.in

Published online: 24 July 2024

© 2024 IACS

core area and diameter of the fiber, thus demonstrating better analyte sensing [6].

The two key important guiding attributes for sensing and chemical sensing are confinement loss and relative sensitivity. The refractive index significantly impacts the Relative Sensitivity; a higher refractive index contrast between the core and cladding ensures the improved light confinement, resulting in increased sensitivity to changes in the surrounding environment. Birefringence of PCF is another remarkable factor taken into consideration, high value of birefringence is inherently preferred. Besides the mentioned characteristics, effective mode area (EMA), numerical aperture (NA), and effective material loss (EML) are also examined to evaluate the efficiency of the sensor.

Due to its unique geometrical structure, PCF-based devices hold an advantage over the ones based on conventional fiber. One can easily elevate the sensing of a PCF by controlling its various design parameters. The design flexibilities make PCF favourable over conventional fibers [7]. By altering the diameter and pitch of holes in the cladding and core region, we can enhance the field interaction with the analyte in the core region of the fiber. The hollow core has a prominent core volume which provides significantly better sensing than the porous-core photonic crystal fibers [8]. A greater core area enhances the interaction of light and the analyte in the core, thus helping in the sensing application.

A variety of distinct biosensors have been introduced in the last two decades for chemical sensing. The innovations have paved a way for design of PCF-based sensors operating in different wavelength regimes. Designs with different core-structure such as porous-core octagonal PCF [9], quasi-PCF [10, 11], and dual-diamond ring PCF [12] was proposed for chemical sensing in terahertz and infrared spectrum. Chemical sensors based on Surface plasmon resonance is another effective way to obtain the sensitivity for a broad range of unknown analytes. Phase matching is a vital condition necessary for SPR, when the frequencies of light photons and electron of the plasmonic material match we observe a sharp loss peak, the fluctuation in the peak is used to identify analytes [13]. A dual-core refractive index fiber is also used for sensing application, the resonant coupling is observed between the two cores, the transmittance of the fiber changes significantly with variations in the refractive index of the surrounding medium [14].

In the past few years, different materials such as Topas [15], Zeonex [16], Teflon [17] and Silicon [18] have been used as background material for PCF based sensors. Material selection is an important aspect, and we chose it based on its distinctive properties which fits the criteria of our study. Topas and Zeonex exhibits similar properties such as constant refractive index (RI) of approximately 1.53 within THz range [19], low material loss, low dispersion, high transparency window, high transition temperatures etc. In our proposed design, we have chosen Zeonex as the bulk material. Zeonex is chosen as the background material over TOPAS and silicon due to its low absorption loss in terahertz regime (0.1–10 THz). The bulk material loss of Zeonex is 0.2 cm^{-1} [20] and the refractive index remains almost constant in the THz range. Zeonex also displays comparatively greater chemical resistivity and higher glass transition temperatures which is an essential factor for the fabrication of fiber [21].

In the past several years, many studies were conducted for the development of PCF based chemical sensors in THz frequency band. In 2023, Hossain et al. suggested a decagonal shaped PCF-based chemical sensor with hexahedron core, using Topas as the bulk material. The relative sensitivity for ethanol, benzene, and water turned out to be 92.55%, 94.65%, 90.30%, respectively. The confinement loss was measure to be 5.81×10^{-08} dB/m for ethanol, 6.01×10^{-08} dB/m for benzene, and 5.84×10^{-08} dB/m for water [6]. In the same year, Hossen et al. [22] proposed another hexahedron-shaped core, with Zeonex as the bulk material for sensing three distinct alcohols. Relative sensitivity and confinement loss for ethanol was reported as 89.50% and 5.45×10^{-08} dB/m, respectively. Ferdous et al. [23] presented an octagonal structure with a square-core. On simulation relative sensitivity for three substance water, ethanol and benzene were 90.65%, 91.93%, and 93.06%, at 1.3 THz. The confinement loss for the same substance were 2.12×10^{-13} dB/m, 2.00×10^{-13} dB/m, and 1.88×10^{-13} dB/m, respectively. In 2021, Sen et al. [24] designed a decagonal PCF with rotated-hexa core and by using Zeonex as the background material. For ethanol, water, benzene the fiber obtained values 78.56%, 79.76% and 77.51% for relative sensitivity, and for confinement loss values were 5.80×10^{-08} dB/m, 6.02×10^{-08} dB/m, and 5.74×10^{-08} dB/m, respectively. In 2020, Iqbal et al. [25] suggested a hexagonal geometry PCF using rectangular air holes for the classification and detection of alcohols. At a monitoring frequency of 2 THz, relative sensitivity for methanol was 88%, for ethanol was 91%, for propanol 92.30%, and for water it was 90.03%. Confinement loss at the optimum frequency for methanol, ethanol, propanol, and water was 0 cm^{-1}, 1.515×10^{-14} cm^{-1}, 5.935×10^{-15} cm^{-1}, and 4.045×10^{-15} cm^{-1}, respectively. In 2019, Sen et al. [26] proposed a circular design with a hexagonal core and used Silicon as the background material. The design was evaluated for ethanol, benzene, and water. At 1 THz, relative sensitivity for ethanol, benzene, and water were 76.44%, 77.16% and 73.20%, respectively. Additionally, confinement loss for the chemicals were 2.33×10^{-03} dB/m, 3.07×10^{-06} dB/m, and 2.84×10^{-02} dB/m. Recently, Kundu et al. proposed a

terahertz PCF sensor for chemical identification and achieved 95.21% for benzene and 94.67% for ethanol and the corresponding confinement loss for benzene and ethanol are 4.674 × 10^{-3} dB/m and 1.378 × 10^{-2} dB/m [27].

The following paper presents a rectangular shaped air hole based PCF chemical sensor model. The effectiveness of the design is assessed on various optical parameters. The optical parameters are calculated for four different analytes- methanol, ethanol, benzene, and water. The model is simulated in the THz range from 0.5 THz to 1.5 THz for evaluation.

2. Design methodology and fabrication possibilities

To study the performance of the proposed PCF sensor, the PCF model is designed using COMSOL Multiphysics software. Figure 1 represents the transverse section view of the proposed sensor model. We have considered methanol, water, ethanol, and benzene as our analytes which are to be infiltrated in the core for sensing purposes. Figure 2 represents the electric field distribution of the propagating fundamental mode in the core region at 1.3 THz. The hollow core photonic crystal fiber offers low effective material loss (EML) and provides higher sensitivity. In this model, the purple-coloured central rectangular hole acts as the core of the fiber in which different analytes are to be filled and the cladding region consists of 16 rectangular air holes which help in guiding the light to propagate through the core region. The total fiber diameter is 2.50 mm and the cladding region is adjoined with a cylindrical layer called perfectly matched layer (PML) to avoid the effect of back reflections into the study area by absorbing the outgoing waves. Generally, the thickness of PML is kept between 2 and 10% of the total fiber diameter [28]. In our simulation, we considered the thickness of PML to be 9% of the fiber

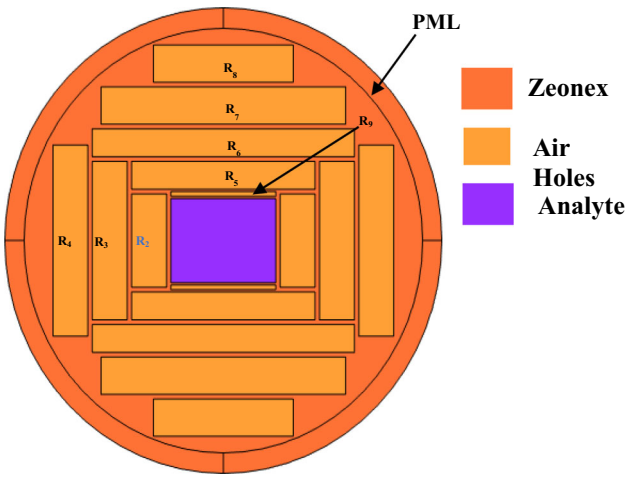


Fig. 1 Proposed structure design

diameter which ensures convergence of the result. The FEM method is executed by using physics-controlled mesh. To get better simulation results we have chosen finer mesh elements having 8742 triangular elements with 1438 edge elements and 76 vertex elements. The height and width of the core are 600 μm and 450 μm, respectively. In cladding region, six rectangles are positioned on left and right sides of the core with the same width but different heights. Identical rectangles are denoted by the same variables. In that way, the nearest rectangles to the core are denoted by R2 with height and width of 500 μm and 200 μm respectively. Similarly, the next nearest rectangles are denoted as R3 and R4 with height and width of 850 μm and 200 μm for R3 and 1025 μm and 200 μm for R4. In the vertical direction, the rectangles above the core region namely R5, R6, R7 and R8 have heights and widths of 150 μm and 1050 μm, 150 μm and 1500 μm, 200 μm and 1400 μm, 200 μm and 800 μm respectively. The pitch in the horizontal direction and vertical direction is kept constant to be 25 μm. Two similar rectangular air holes are also arranged just above and below the core to confine the propagating electric field more effectively in the core region by creating the index contrast. They are denoted as R9 and have height and width of 25 μm and 600 μm. The dimensions of the core must be kept as large as possible to get higher values of core power fraction. But keeping in mind the feasibility and fabrication complexities, we cannot choose the dimensions as per free will. In our proposed model, while simulation we have kept the width and height of the rectangular holes to be the same and varied the dimensions of the rectangular core region by 2% to experience the change in sensing performance of the model. After that, we stated the height and width of the core region as optimum (OPT). In recent times, various PCF fabrication techniques and methods have been used to fabricate different PCF geometry structures. For the fabrication of circular air holes, stacking and sol–gel methods are widely used [29–31]. Furthermore, for the fabrication of asymmetric PCF structure extrusion method and 3-D printing techniques [32, 33] are used. Using the extrusion method rectangular and spider-web-shaped PCFs have already been realized which opens up the possibility of fabrication of the proposed design structure [32]. Hence, the fabrication of the proposed PCF structure is attainable by the subsisting modern fabrication techniques.

3. Numerical analysis

The following study is conducted for four different analytes methanol, ethanol, benzene, and water. To assess the performance of the proposed chemical sensor, different analytes are separately injected into the rectangular core

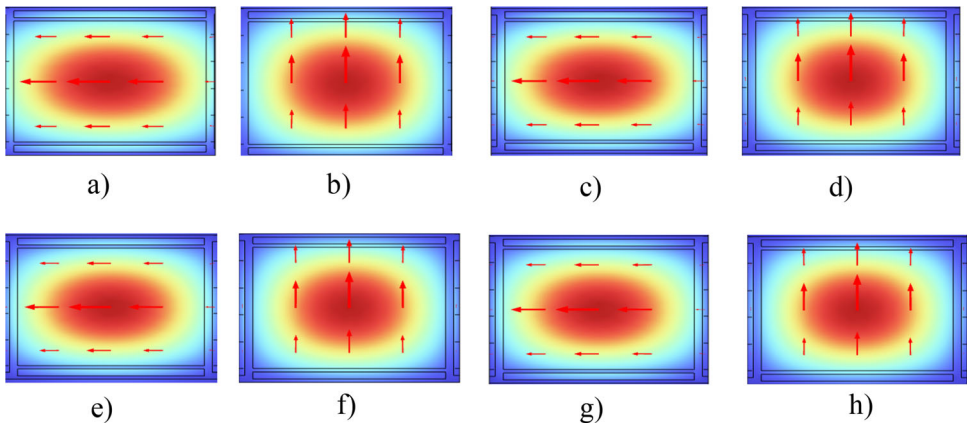


Fig. 2 Mode field distribution of fundamental mode for (**a**) Methanol, x-pol (**b**) Methanol, y-pol (**c**) Water, x-pol (**d**) Water, y-pol (**e**) Ethanol, x-pol (**f**) Ethanol, y-pol (**g**) Benzene, x-pol (**h**) Benzene, y-pol

and are studied. The effectiveness of the sensor depends on various optical parameters which include, Relative sensitivity, Power fraction, Confinement loss, Effective mode area, Effective material loss, Numerical Aperture, Birefringence etc.

When light propagates through a fiber it experiences an average refractive index which is determined by the design of the PCF. This mean refractive index is referred to as the effective refractive index.

The key element for analysing the performance of a sensor is the Relative Sensitivity (RS), which is calculated by the following equation [34]:

$$RS = \frac{n_a}{n_{eff}} \times P \tag{1}$$

where n_a represents the refractive index of the analytes which is 1.317 for Methanol, 1.354 for ethanol, 1.330 for water, and 1.366 for benzene, n_{eff} stands for effective-mode index and P for power fraction.

In PCF-based chemical sensors, power fraction can be defined as the amount of propagating light that interacts with the analyte present within the core region. Power fraction (PF) is calculated using the equation given below [35]:

$$P = \frac{sample \int R\left(E_x H_y - E_y H_x\right) dxdy}{whole \int R\left(E_x H_y - E_y H_x\right) dxdy} \tag{2}$$

In the above equation, we perform integration to calculate the light propagated within a particular analyte concerning the cross section of the fiber. Here, Ex, Ey represents the electric field components and Hx, Hy represents the magnetic field components.

Another important optical parameter which helps in the analysis of a PCF-based sensor is Confinement Loss. Confinement loss (CL) is the propagation loss due to leakage which determines the light confining ability of any PCF. The lower the confinement loss, better will be the confining ability which is a pre-requisite for any PCF based

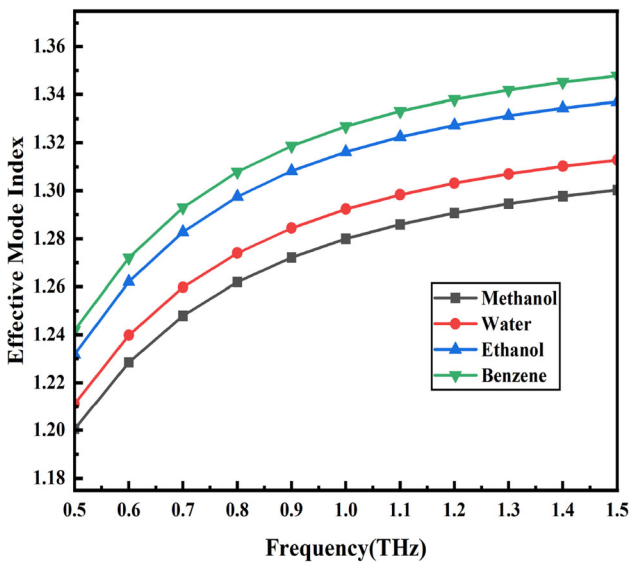


Fig. 3 Effective mode index variation (x-pol)

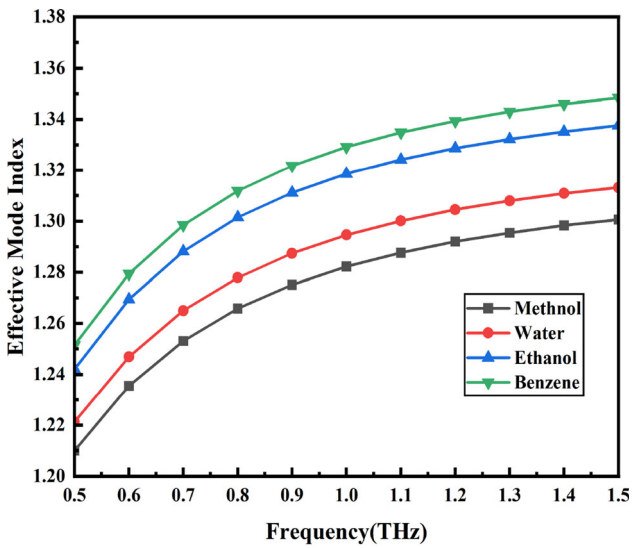


Fig. 4 Effective mode index variation (y-pol)

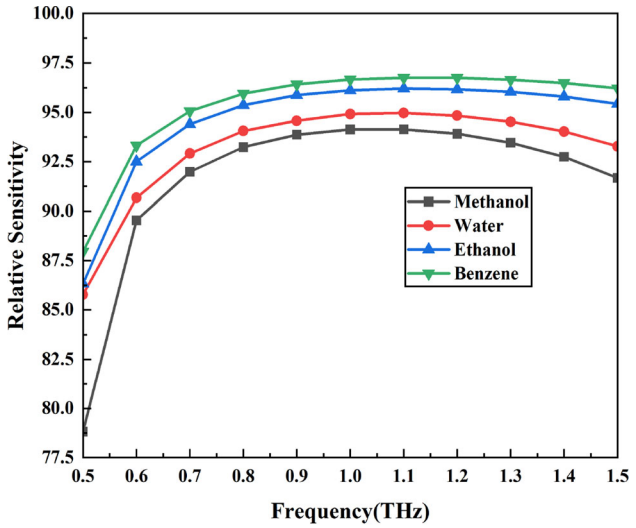


Fig. 5 Relative Sensitivity for OPT-2% profile (x-pol)

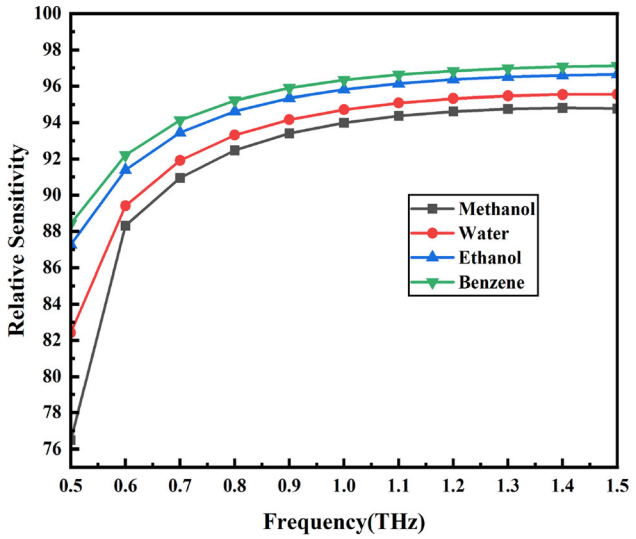


Fig. 6 Relative Sensitivity for OPT-2% profile (y-pol)

sensor model. CL depends on the imaginary part of effective mode index and is given as [24]:

$$C.L = 8.868 \times k_o I_m\left(n_{eff}\right) dB/m \tag{3}$$

Here, k_o is the propagation constant and $I_m\left(n_{eff}\right)$ is the imaginary part of the effective refractive index of fundamental mode.

Effective mode area (EMA) is also an important optical parameter in any PCF. PCFs with low effective mode area can be useful in non-linear optics and PCFs with large effective mode area can find potential applications in laser communications as well as in optical and electronic devices. It gives the idea about the amount of light propagating in the cladding region of the fiber and in the core region where the analyte is to be filled. In a PCF-based sensor

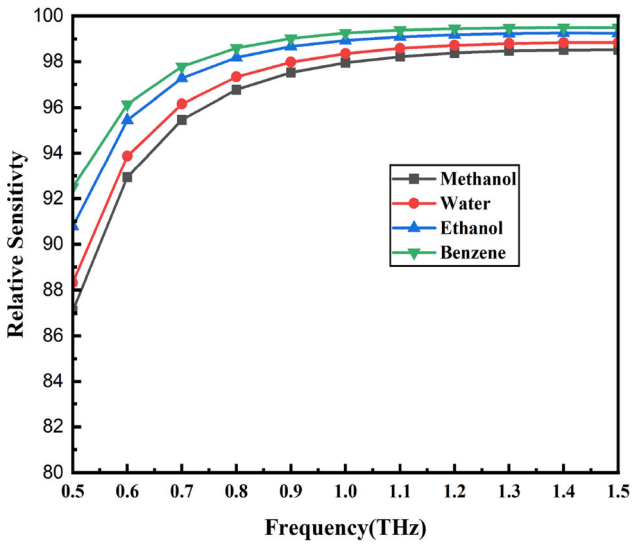


Fig. 7 Relative Sensitivity for OPT+2%profile (x-pol)

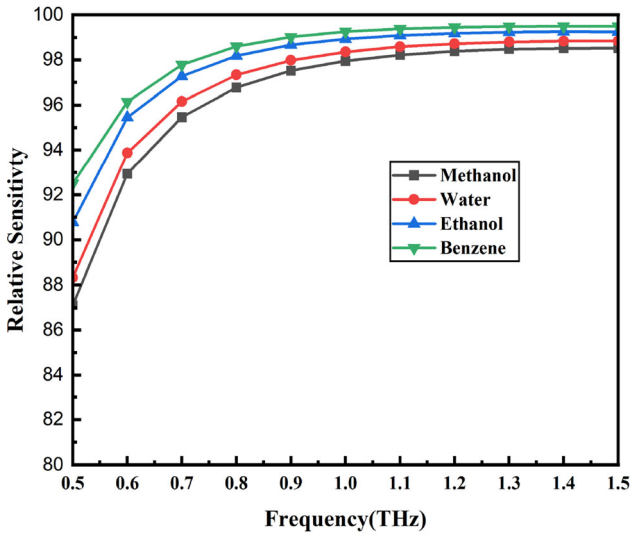


Fig. 8 Relative Sensitivity for OPT+2% profile (y-pol)

model, the high effective mode area indicates high relative sensitivity and low confinement loss. EMA which is usually referred as A_{eff} is calculated using the equation given below [36]:

$$A_{eff} = \frac{\left(\iint |E|^2 dxdy\right)^2}{\iint |E|^4 dxdy} \tag{4}$$

Here, |E| represents the electric field distribution of the suggested fiber sensor.

Effective material loss (EML) is another loss which comes into play during the propagation of light due to absorption by the material of the fiber. EML can be calculated as [35]:

Table 1 Comparative study of relative sensitivity of all four analytes for y polarization mode

Analyte	Profile	Operating frequency (THz)	R.S (%)
Methanol	OPT − 2%	1.3	94.75
	OPT	1.3	96.85
	OPT + 2%	1.3	98.35
Water	OPT − 2%	1.3	95.46
	OPT	1.3	97.32
	OPT + 2%	1.3	98.66
Ethanol	OPT − 2%	1.3	96.50
	OPT	1.3	97.99
	OPT + 2%	1.3	99.14
Benzene	OPT − 2%	1.3	96.97
	OPT	1.3	98.33
	OPT + 2%	1.3	99.32

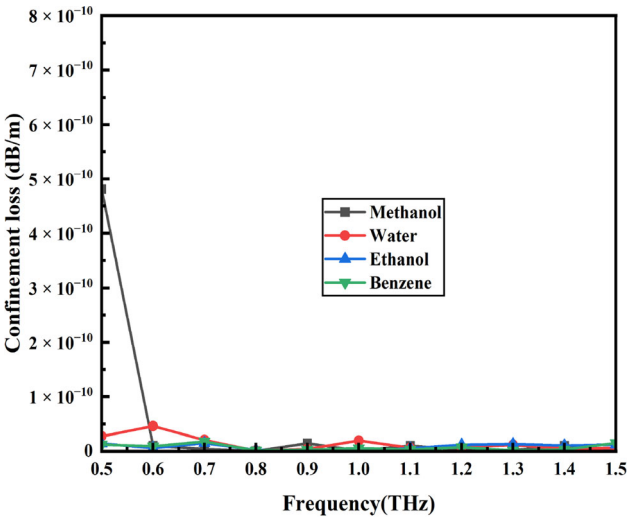


Fig. 10 Confinement loss as a function of frequency

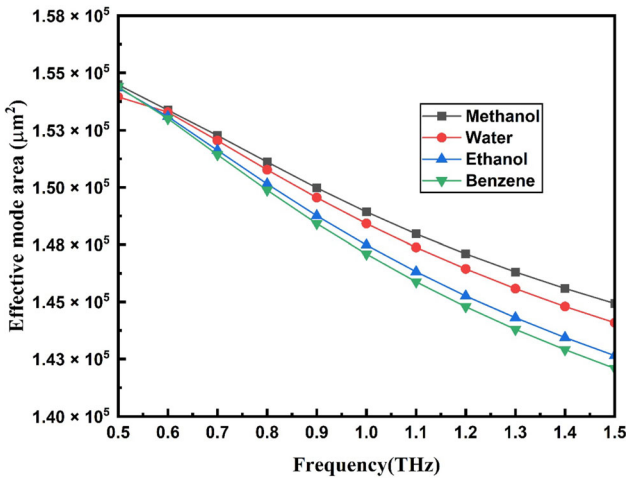


Fig. 9 Effective mode area as a function of frequency

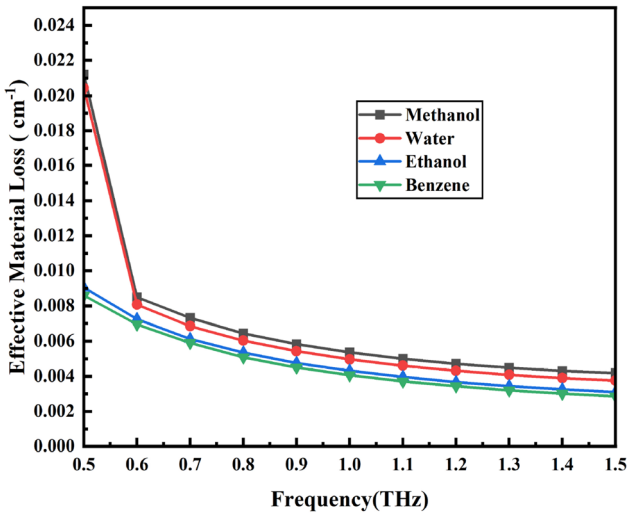


Fig. 11 Effective material loss as a function of frequency

$$\alpha_{eff} = \sqrt{\frac{\varepsilon_o}{\mu_o}}\left(\frac{\int n_{mat}|E|^2\alpha_{mat}dA}{|\int P_z dA|}\right) \quad (5)$$

In the above equation, ε_o is the permittivity in free space and μ_o is the permeability in the free space. n_{mat} represents the refractive index and αmat represents the bulk absorption loss of the Zeonex. P_z represents the Poynting vector in the z direction.

Numerical Aperture is another important characteristic which is significant to check the efficiency of any PCF-based sensor model. Generally, a large numerical aperture is preferred for better sensing results as the higher the NA of the fiber higher the light-gathering capacity which further improves the field interaction with the analyte [37].

$$NA = \left(1 + \frac{\pi A_{eff} f^2}{c^2}\right)^{-\frac{1}{2}} \quad (6)$$

In terahertz communication systems and sensing applications, it is necessary to maintain a high polarization state. Birefringence is the property which helps in maintaining the polarization state of a fiber. Birefringence is defined as the difference in effective refractive indices of the fundamental mode of the fiber in x and y polarizations. It can be calculated by [35]:

$$B = |n_x - n_y| \quad (7)$$

n_x and n_y here represent the effective mode indices of *x*-polarization and *y*-polarization, respectively.

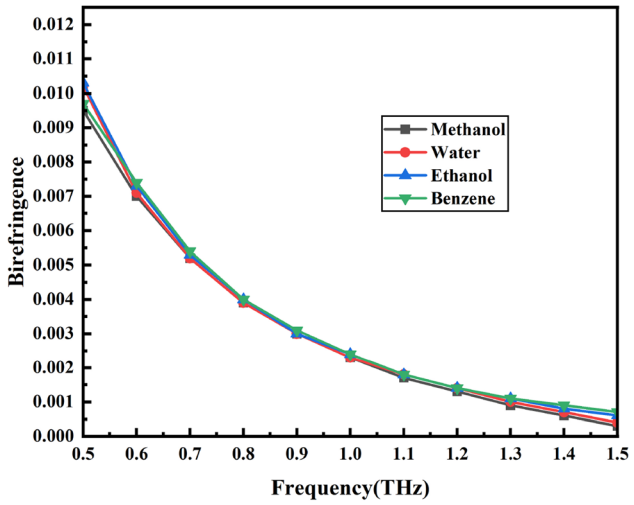


Fig. 12 Birefringence as a function of frequency

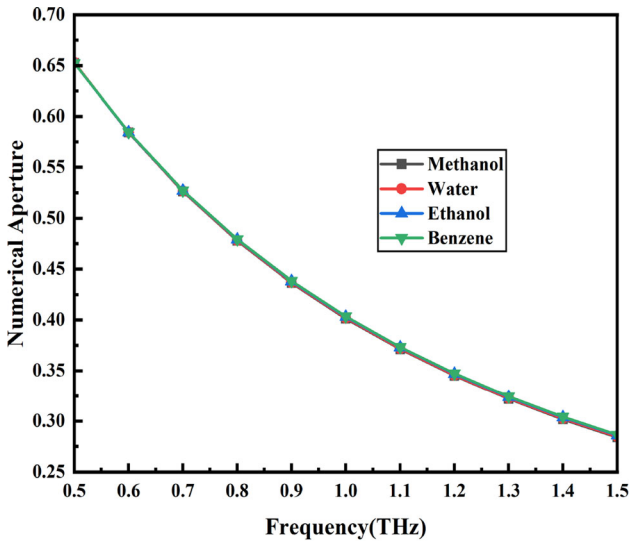


Fig. 13 Numerical aperture as a function of frequency

4. Results and discussion

The proposed sensor model is simulated using COMSOL Multiphysics software and all the optical characteristics are calculated using the same tool. We get the results in the frequency range of 0.5–1.5 THz, and the declared optimum frequency is 1.3 THz. We have calculated the Relative Sensitivity (for both *x* and *y*- polarization modes), PF, CL, A_{eff}, EML, NA and birefringence.

We have first analysed the effect of frequency of the effective-mode-index of the propagating mode for all the analytes for the both polarization and shown in Figs. 3 and 4 respectively. As expected, an increase in the effective mode index with frequency is observed for all analytes for both polarizations. As the refractive index of benzene is highest among ethanol, water and methanol hence the effective-mode-index of benzene is highest followed by ethanol, water and methanol at a particular frequency.

As mentioned earlier, we have considered three different core design parameters (OPT −2%, OPT, OPT + 2%). The relative sensitivities of methanol, water, ethanol and benzene as a function of frequency are shown graphically in Figs. 5, 6, 7 and 8. In these figures, one can observe that RS increases with frequency. Confinement of the modal field increases with frequency and hence increases in the power fraction, *P*, thus increasing relative sensitivity. The relative sensitivity also depends proportionally on the RI of the analyte being used, as with high-index, the confinement of light is high. As a result, the relative sensitivity for benzene is highest among all analytes. Initially, the values of relative sensitivities for x polarization mode is greater than that of y polarization mode as the area of the proposed structure's core is higher in x-direction but after 1.2 THz the y polarization mode has higher relative sensitivities as light confinement becomes effective due to the presence of two nearly arranged holes in the y-direction. For OPT-2% profile, the relative sensitivities of methanol, water, ethanol and benzene are 93.46%, 94.52%, 96.03%, and 96.66% in x-polarization mode and are 94.75%, 95.46%, 96.50%, and 96.97%, respectively in y- polarization mode. For OPT profile the relative sensitivities for methanol, water, ethanol, and benzene are 96.57%, 97.15%, 97.96%, and 98.35% in x-polarization mode and are 96.85%, 97.32%, 97.99%, and 98.33% in y-polarization mode at operating frequency of 1.3 THz. For OPT + 2% profile case, the calculated relative sensitivity for benzene reaches 99.34% but as we increase the length and width of the core by 2%, the chances during fabrication that the core and cladding holes get merged increase, therefore we chose the OPT profile case as a compromise. The comparative study of the relative sensitivity of all four analytes for the y-polarization mode is shown in Table 1.

Figure 9 shows the variation of Effective mode area as a function of frequency. It can be observed that the EMA of the OPT profile decreases with the increase in frequency. High operation frequency causes tight confinement of light in the core region, thus reduces the mode area of the propagating mode. At operating frequency of 1.3 THz, the values of effective mode area are 1.46×10^5 μm^2 for methanol, 1.45×10^5 μm^2 for water, 1.44×10^5 μm^2 for ethanol and 1.43×10^5 μm^2, for benzene.

Confinement loss for the OPT. profile is shown graphically as a function of frequencies in Fig. 10. As the frequency increases the value of confinement loss decreases because of the effective confinement of light at larger frequencies. Confinement loss for benzene is lowest among all other analytes as the R.I of the benzene is highest,

Table 2 Comparative study of optical parameters among previous works and proposed work

References	Operating region (THz)	Design (core)	Analyte	Relative sensitivity (%)	Confinement loss (dB/m)	Material loss (cm^{-1})
[22]	1	Hexahedron	Ethanol	90.50	5.45×10^{-08}	–
[23]	1.3	Square	Water	90.65	2.12×10^{-13}	0.0136
			Ethanol	91.93	2.0×10^{-13}	0.0146
			Benzene	93.06	1.88×10^{-13}	0.0156
[24]	1	Rotated-hexagonal	Water	77.51	5.74×10^{-08}	–
			Ethanol	78.56	6.02×10^{-08}	–
			Benzene	79.76	5.80×10^{-08}	–
[25]	2	Rectangular	Methanol	89.29	0 cm^{-1}	0.0056
			Water	90.03	1.5×10^{-15}	0.0059
			Ethanol	91.30	$5.935 \times 10^{-}$	0.0064
[26]	1	Rotated-hexagonal	Water	73.20	2.84×10^{-02}	–
			Ethanol	76.44	2.33×10^{-03}	
			Benzene	77.16	3.07×10^{-05}	
[27]	-	Hexagonal	Ethanol	94.67	1.378×10^{-2}	0.0083
			Benzene	95.21	4.674×10^{-3}	0.0078
This paper	1.3	Rectangular	Methanol	96.85	2.22×10^{-12}	0.0044
			Water	97.32	1.16×10^{-11}	0.0040
			Ethanol	97.99	1.34×10^{-11}	0.0034
			Benzene	98.33	1.30×10^{-12}	0.0032

reason being the same that light gets more confined with analytes having greater refractive indices. As observed from the graph the values below 10^{-11} dB/m are unstable reason being the values are too small and are comparable to the numerical errors introduced during the simulations [3]. The values of confinement loss at 1.3 THz are is 2.22×10^{-12} dB/m for methanol, 1.16×10^{-11} dB/m for water, 1.34×10^{-11} dB/m for ethanol and, 1.30×10^{-12} dB/m for benzene.

Effective material loss is shown graphically as a function of frequencies in Fig. 11. For a PCF based sensor low EML is preferred as high values of material loss limits the sensor performance. At operating frequency region of 1.3 THz, the proposed sensor offers low values of EML which are 0.0044 cm^{-1} for methanol, 0.0040 cm^{-1} water, 0.0034 cm^{-1} for ethanol and 0.0032 cm^{-1}, for benzene.

Figure 12 illustrates the birefringence as a function of frequency. Due to the asymmetry between the core and cladding air holes of the fiber, birefringence will also be present in our designed structure. Birefringence decreases as the frequency increases because the difference between the effective indices of x and y polarizations decreases with the increase in frequency. The optimum profile offers birefringence around 0.0009 for methanol, 0.0010 for water, 0.0011 for ethanol and 0.0011, for benzene at operating frequency of 1.3 THz.

Figure 13 illustrates the variation of numerical aperture as a function of frequencies. As it is logical that NA decreases with the increase in frequency which is verified by the graph in Fig. 13. The obtained values of NA at 1.3 THz are 0.32 for methanol, 0.40 for water, 0.32 for ethanol and 0.32 for benzene. To get an overall idea about the performance of proposed PCF based chemical sensor model, the comparative result of the sensing properties are tabulated and represented in Table 2.

5. Conclusion

In conclusion, a PCF based chemical sensor with rectangular-core model is presented for its sensing application in THz region. Using Zeonex as the background material, various optical parameters of the proposed sensor model have been examined over the range of 0.5–1.5 THz. All the numerical calculations are performed by PML boundary condition using FEM based COMSOL Multiphysics software. The optimum profile of the proposed sensor model exhibits high relative sensitivities of 96.85%, 97.32%, 97.99%, 98.33%, and low values of confinement loss as 2.22×10^{-12} dB/m, 1.16×10^{-11} dB/m, 1.34×10^{-11} dB/m, 1.30×10^{-12} dB/m for methanol (1.317), water (1.33), ethanol (1.354) and benzene (1.366) respectively at 1.3

THz operating frequency. Further, effective material loss for the model also comes out to be very low which makes our proposed sensor model worthy of proficient sensing. So, the proposed PCF sensor model should be helpful in industrial areas for chemical sensing, food and biosesning applications. Furthermore, with the advancement in modern fabrication techniques, the fabrication of the proposed PCF is attainable.

Acknowledgements This work was supported by the Delhi Technological University, Bawana Road, New Delhi, Delhi 110042, India.

References

[1] M Rachana, I Charles, S Swarnakar, S V Krishna and S Kumar *Opt. Fiber Technol.* **74** 103085 (2022)
[2] A Bulbul, H Rahaman, S Biswas, M B Hossain and A Nahid *Sens. Bio Sens. Res.* **30** 100388 (2020)
[3] T Yang, L Zhang, Y Shi, S Liu and Y Dong *Sensors* **21** 1799 (2021)
[4] H Xin and M Liang *Proc. IEEE 105* **4** 737 (2017)
[5] T Yang, C Ding, R W Ziolkowski and Y J Guo *Opt. Express* **26** 17264 (2018)
[6] M S Hossain, R Hossen, S T Alvi, S Sen, M A Amin and M M Hossain *Phys. Open* **17** 100168 (2023)
[7] D J J Hu and H P Ho *Adv. Opt. Photon.* **9** 257 (2017)
[8] M R Islam, F A Mou, M M Rahman and M I Hassan Bhuiyan *Int. J. Mod. Phy. B* **34** 2050259 (2020)
[9] B K Paul, M S Islam, K Ahmed and S Assaduzaman. *Photon. Sens.* **7** 123 (2017)
[10] B K Paul, K Ahmed, D Vigneswaran, F Ahmed, S Roy and D Abbott *IEEE Sens. J.* **18** 9948 (2018)
[11] B K Paul, K Ahmed, D Vigneswaran, S Sen and M S Islam. *Opt. Quant. Electrons.* **51** 238 (2019)
[12] B K Paul, K Ahmed, H J E Khozondar, et al. *J. Comput. Electron.* **19** 1288 (2020)
[13] S Sarker, D Vigneswaran and F Studnička *Front. Phys.* **11** 1164255 (2023)
[14] K Kamakshi, V Rastogi and A Kumar *Opt. Fiber Technol.* **19** 325 (2013)
[15] V S Chaudhary and D Kumar *Optik* **223** 223 (2020)
[16] S Sen, M A Shafi, A S Sikder, M S Hossain and M M Azad *Sens. Bio Sens. Res.* **31** 100393 (2021)
[17] R Senthil, U Anand and P Krishnan *Appl. Phys. A* **127** 1 (2021)
[18] M S Hossain and S Sen *Silicon 13* **11** 3879 (2021)
[19] M S Islam, J Sultana, A Dinovitser, M Faisal, M R Islam, B W Ng and D Abbott *Appl. Opt.* **57** 666 (2018)
[20] M S Islam, J Sultana, A Dinovitser, B W Ng and D Abbott *Opt. Commun.* **413** 242 (2018)
[21] G Woyessa, A Fasano C Markos, A Stefani, H K Rasmussen and O Bang *Opt. Mater. Express* **7** 286 (2017)
[22] R Hossen, M S Hossain, S Ahmed, M Sayduzzaman, M Sultana, R Laizu Naima and S Sen *Phys. Open* **17** 100192 (2023)
[23] A H M I Ferdous, A H M Iftekharul, B Akter, A Priya, L M Leo, R T Prabu, B N Sathi, D Kundu et al. *Plasmonics* **19** 1257 (2023)
[24] S Sen, M A Shafi, A S Sikder, M S Hossain and M M Azad *Sens. Bio Sens. Res.* **31** 100393 (2021)
[25] F Iqbal, S Biswas, A A M Bulbul, H Rahaman, M B Hossain, M E Rahaman and M A Awal *Sens. Bio Sens. Res.* **30** 100384 (2020)
[26] S Sen and K Ahmed *SN Appl. Sci.* **1** 1215 (2019)
[27] D Kundu, M O Faruk, K Khandakar, A H M L Ferdous, M M Bani and K S Noor *Int. Commun. Heat Mass Transf.* **156** 107615 (2024)
[28] M E Rahaman, H S Mondal, M B. Hossain, M M. Hossain, M S Ahsan and R Saha *SN Appl. Sci.* **2** 1 (2020)
[29] P S J Russell *J. Lightw. Technol.* **24** 4729 (2006)
[30] K Nielsen, K R Henrik, A J L Adam, P Planken, O Bang and P Jepsen *Opt. Express* **17** 8592 (2009)
[31] H Hamzaoui, Y Ouerdane, L Bigot, G Bouwmans, B Capoen, A Boukenter, S Girard and M Bouazaoui *Opt. Express* **20** 29751 (2012)
[32] S Atakaramians, S Afshar, H E Heidepriem, M Nagel, B M Fischer, D Abbott and T M Monro *Opt. Express* **17** 14053 (2009)
[33] H E Heidepriem, J Schuppich, A Dowler, L Marques and T M Monro *Opt. Mater. Express* **4** 1494 (2014)
[34] M E Rahaman, R H Jibon, H S Mondal, Md B Hossain, A A Bulbul and R Saha *Sens. Bio Sens. Res.* **32** 100422 (2021)
[35] M B Hossain, E Podder, A A Bulbul and H S Mondal *Opt. Fiber Technol.* **54** 102102 (2020)
[36] S Yadav, P Lohia and D K Dwivedi *J. Opt.* **53** 2642 (2023)
[37] M T Islam, M G Moctader, K Ahmed and S Chowdhury *Photon. Sens.* **8** 263 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

2024 International Conference on Advances in Modern Age Technologies for Health and Engineering Science (AMATHE) | 979-8-3503-7156-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/AMATHE61652.2024.10582218

Document Information Retrieval with Deep Learning

Abhilasha Sharma
Department of Software Engineering
Delhi Technological University,
Delhi, India
abhi16.sharma@gmail.com

Aparna Arya
Department of Software Engineering
Delhi Technological University
Delhi, India
aparna_2k22swe03@dtu.ac.in

***Abstract*— Retrieving relevant medical data promptly is essential for both improving patient care and scientific discoveries. To compare the performance of Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Deep Neural Networks (DNNs), and a unique fusion approach called Random Multimodel Deep Learning (RMDL), this study addresses the application of deep learning models in medical document retrieval. The findings of this research emphasize the value of RMDL's multi-model fusion and imply that integrating several learning modalities can have positive effects in the medical field. improved document retrieval systems have the potential to provide medical personnel with faster access to pertinent information, which could result in more precise diagnoses, more focused treatment regimens, and ultimately improved patient outcomes. Improved retrieval also makes it easier to collect information for medical research in an efficient manner, which speeds up breakthroughs across a range of domains and aids in the creation of individualized medicine strategies. Ultimately, the most effective model is assessed using performance evaluation parameters.**

Keywords- Medical data, CNN, RNN, DNN, RMDL fusion

I. INTRODUCTION

Retrieving document information is the process of comparing a group of free-text records to a user-specified query. Any kind of primarily unstructured material, including newspaper articles, real estate records, and manual paragraphs, could be included in these records. User queries might be as short as a few words or as long as many sentences that fully describe what information is needed. Sometimes, Document Information Retrieval (DIR) is referred to as text retrieval, or as a subset of it. Within the field of information retrieval, text retrieval deals with information that is predominantly stored as text. Since text retrieval is the foundation of all Internet search engines, it is an important field of study today.

The use of neural network architectures such as CNNs, RNNs, and transformers in the analysis and extraction of data from text documents is part of Document Information Retrieval (DIR) with Deep Learning [1]. These models are learning how to interpret the content of documents in a sophisticated way to improve tasks like document classification, grouping, and relevance ranking. By analyzing the semantic meaning and context of documents, deep neural networks can improve document retrieval systems' accuracy and efficiency.

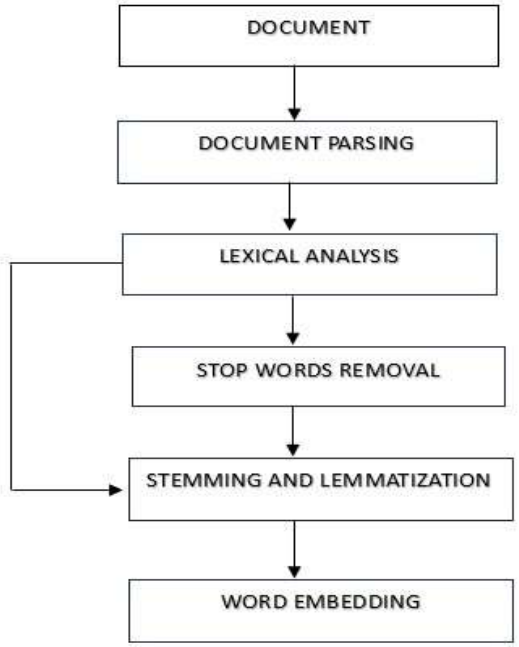


Fig. 1. Document Information Retrieval

A CORD 19 [2] dataset with 1,00,000 records is first collected. Importing necessary libraries such as NumPy, Pandas, and Matplotlib allows for the manipulation and analysis of data, the support of multi-dimensional arrays and mathematical functions by NumPy, and data visualization by Matplotlib. After loading the dataset into a Pandas data frame, actions for data cleaning are performed to address missing, null, and duplicate values as well as to eliminate superfluous symbols like commas, question marks, and punctuation.

During the pre-processing phase, the document is divided into discrete parts using document parsing. Subsequently, the significant pieces are tokenized and transmitted for Parts of Speech Tagging through lexical analysis. By eliminating affixes and locating roots, the stemming and lemmatization processes improve words. Last but not least, Word Embedding converts every document into a fixed-length vector that represents the frequency of words.

Using Principal Component Analysis, or PCA, dimensionality reduction is carried out. The technique known as *"dimensionality reduction"* essentially aids in improving

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:24 UTC from IEEE Xplore. Restrictions apply.

our understanding of our facts on a macro level. Our dataset's features are reduced to just the most crucial elements by doing this. A powerful yet straightforward method for reducing dimensionality is principal component analysis or PCA. It allows to immediately reduce the number of feature variables, focusing on the most crucial aspects and reducing the number of computations required. The Doc2Vec [3] approach is then used to extract the data. Regardless of a document's length, doc2vec aims to generate a numerical representation of it. A generalization of the word2vec method, Doc2vec is an NLP tool for expressing documents as a vector.

The K-means algorithm [4], a widely used method for exploratory data analysis, is used to cluster data. It seeks to locate subsets in the data where data points in one cluster are comparable to each other but data points in other clusters are not. By reducing the squared distance between data points and cluster centroids, K-means divides the dataset into pre-defined clusters and encourages both intra- and inter-cluster dissimilarity.

The organization of this paper is as follows: Section II elaborates on the literature review for Document Retrieval using various NLP algorithms. Section III depicts, in a nutshell, the various deep learning models used in this research paper for document retrieval purposes. Section IV describes the dataset and various parameters based on which describes the dataset and various parameters based on which the best model will be evaluated. Section V gives the conclusion and suggests topics for further research.

II. LITERATURE SURVEY

The technique for creating deep sentence embeddings using long short-term memory (LSTM) networks is presented by Palangi, H., Deng, L.[5]. By calculating the semantic similarity between queries and documents, these embeddings are used to improve information retrieval. Results from experiments show how well LSTM-based embeddings work to increase retrieval relevance and accuracy in NLP tasks.

Neural information retrieval (IR) strategies are reviewed in detail by Zhang, Y., Rahman.[6] The study looks at different neural network models and architectures used for information retrieval tasks like query comprehension and document ranking. In using neural networks to improve information retrieval systems, the article outlines developments, obstacles, and potential paths forward.

The performance of deep learning algorithms for biomedical document classification is assessed in this research by Behera, B., Kumaravelan.[7] The classification of biomedical documents was evaluated using several deep learning models at the 11th International Conference on Advanced Computing, 2019. The study adds to our understanding of how well deep learning methods work for classifying and organizing biological material.

III. DEEP LEARNING MODELS

Progressive outcomes have been attained by deep learning models in NLP application domains. Basic structured deep learning models and Ensemble deep learning models can be used to group all deep learning models for document information retrieval.

The term "Basic Structure DL models" refers to deep learning models with their own basic deep learning structure, such as CNN, DNN, and RNN. Ensemble Deep Learning models are those types of deep learning models that are constructed by integrating fundamental deep learning models, such as RMDL. The following is a brief explanation of those DL methods.

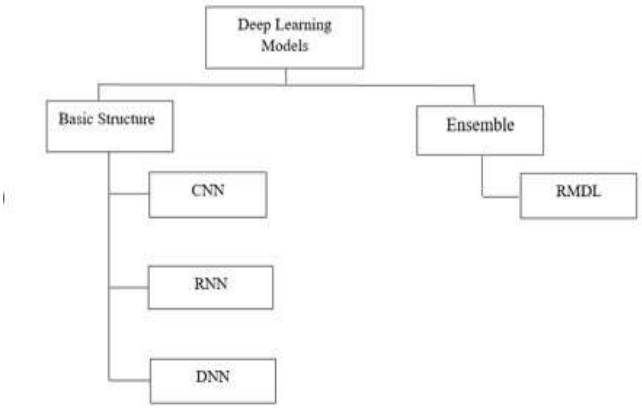


Fig. 2. Segmentation of Deep Learning Models

A. Deep Neural Network

In contrast to MLP, DNN [8] has a large number of hidden layers. Following that, the neural network is trained by the conventional backpropagation process. Nodes in the input layer are equal to the number of characteristics that were extracted. For multiclass classification, SoftMax is employed as the activation function in the output layer and sigmoid and Re-LU are utilized in the hidden layers. The prediction model is employed in the output layer. The number of nodes in the output layer corresponds to the number of classes in the dataset.

B. Convolutional Neural Network

With its several layers, including embedding, convolutional, and pooling layers, CNN [9] offers the best feature extraction techniques. As the input layer, the embedding layer transforms each word from the pre-processed texts into an embedding vector with an embedding dimension. The dimension of the Convolution layer is decreased by the Pooling Layer. After receiving two-dimensional representations of words or features, the Flatten Layer transforms them into vector form that the Fully Connected Layer can use. One or more neural layers that employ activation functions and dropout rates to learn the model are referred to as the fully connected layer or dense layer. Model prediction is done using the Output Layer, whose activation function is SoftMax and whose nodes are equal to the number of classes in the dataset.

C. Recurrent Neural Network

Better semantic analysis of the dataset is made possible by the RNN's [10] recurrent structure, which conveniently takes into account the information from earlier nodes. Every word in the pre-processed texts is transformed into embedding vectors with an embedding dimension by the embedding layer of the RNN model, which functions as the input layer. To learn the model, RNN architectures like GRU or LSTM are employed in conjunction with the dense layer. Nodes equal to the number of classes in the dataset are present in the output layer, which employs SoftMax as its activation function.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:24 UTC from IEEE Xplore. Restrictions apply.

D. Random Multimodel Deep Learning

By using ensembles of different underlying deep learning architectures, such as DNN, RNN, and CNN, RMDL [7,11] improves accuracy while simultaneously solving the problem of identifying the simplest deep learning structure and design

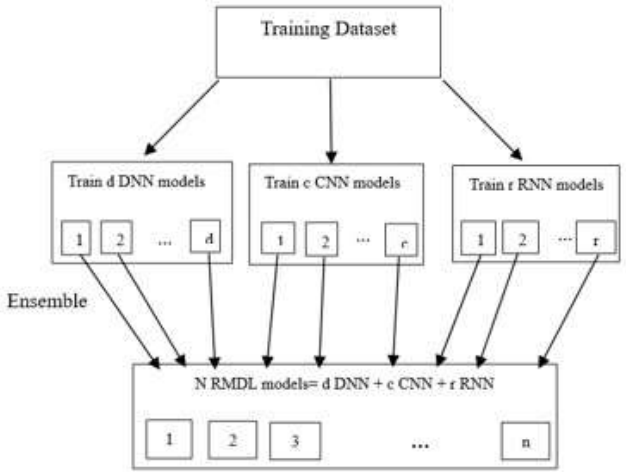


Fig. 3. RMDL Model Overview

A training dataset is given to each of the component DL models, such as DNN, CNN, and RNN, to construct the RMDL model. With the help of the Python-based sk-learn library's TfidfVectorizer [12], the DNN models preprocess the training dataset. Similar to this, Glove pre-trained word embedding and Keras tokenizer are used by the CNN and RNN models to preprocess the training dataset. Create d, c, and r numbers of CNN, RNN, and DNN models from the training dataset during the training phase. Thus, the sum of d, c, and r represents the total number of models in RMDL. After the testing data is loaded into the final neural model, the recovered documents will be evaluated for ranking and relevancy, and any additional features that are needed will be improved or integrated.

IV. RESULTS AND DISCUSSIONS

A. Dataset

The COVID-19 Open Scientific Dataset (CORD-19) [2] was created in reaction to the epidemic by the White House and a partnership of eminent scientific organizations. CORD-19 is a collection of more than 200,000 academic publications about COVID-19, SARS-CoV-2, and similar coronaviruses, of which more than 100,000 are available in full text.

B. Performance Metrics

Four features—true positive, true negative, false positive, and false negative—are used to establish performance measures like accuracy, precision, recall, and F1-score. Performance metrics including accuracy, weighted-average-based precision, recall, and F1-score are utilized in this experiment to assess the classifier's effectiveness. Here is a definition of the performance measures
which more than 100,000 are available in full text.

1. Precision is the percentage of all affirmative cases that were accurately identified.

Precision = True Positives / (True Positives + False Positives) (1)

2. The percentage of accurately recognized positive instances relative to all real positive cases is recall

Recall = True Positives / (True Positives + False Negatives (2)

3. An impartial indicator of model performance is the F1-score, which is the harmonic mean of precision and recall.

F1 Score = 2 * (Precision * Recall) / (Precision + Recall) (3)

C. Results

The results of a comprehensive experiment on various Deep Learning (DL) algorithms, including DNN, RNN, CNN, and RMDL, for the COVID-19 medical dataset CORD 19, are examined in this part. In this work, we integrated the DNN, RNN, and CNN models after using 200,000 words for the embedding layer. 30,000 of the 200,000 queries were utilized for testing the model and 30,000 for training.

TABLE I. Performance of DL Models

PERFORMANCE MEASURES	*DNN*	*RNN*	*CNN*	*RMDL*
Accuracy	87.43	82.65	98.84	98.96
Precision	86.2	83.4	97.1	98.24
Recall	87.12	85.7	98.3	98.31
F1-Score	86.3	84.5	97.7	98.26

We examined the effectiveness of several deep learning models for the CORD 19 dataset in this study. Convolutional Neural Networks (CNNs) had the highest accuracy (98.84%) among the individual models, followed by Recurrent Neural Networks (RNNs, 82.65%), Deep Neural Networks (DNNs, 87.43%), and Random Multimodel Deep Learning (RMDL) models (98.96%). While reaching equivalent accuracy to CNN, RMDL surpassed both DNN and RNN across all evaluation criteria (precision, recall, F1 score). Notably, RMDL outperformed CNN in terms of accuracy (0.13%) and F1 score (0.56%), indicating that it may be able to increase performance in situations where individual models may not be at their best.

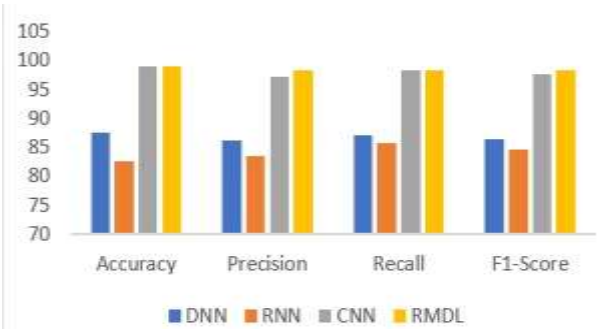


Fig. 4. Comparison of Performance Metrics

V.CONCLUSION

Important insights from our investigation into deep learning models for medical document retrieval are obtained. The new Random Multimodel Deep Learning (RMDL) method proved to be a powerful rival, even if individual CNNs achieved the best standalone accuracy. RMDL drove

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:24 UTC from IEEE Xplore. Restrictions apply.

both DNNs and RNNs in crucial parameters such as precision, recall, and F1 score, even though CNNs and RMDLs exhibited comparably high accuracy. Indeed, RMDL exceeded CNNs in terms of accuracy and F1 score, indicating the potential for outperforming individual models under certain conditions. These outcomes highlight the way RMDL performs in the retrieval of medical documents and opens up new avenues for study. Further research into the distinct design and hyperparameter tuning of RMDL is necessary to completely understand its benefits and drawbacks in a variety of medical subdomains and retrieval activities. Furthermore, adding interpretability strategies to RMDL may improve its clinical relevance by providing insightful information about how it makes decisions.

REFERENCES

[1] R. Zhu, X. Tu, and J. Xiangji Huang, "Deep learning on information retrieval and its applications," Deep Learning for Data Analytics, pp. 125–153, 2020. doi:10.1016/b978-0-12-819764-6.00008-9.

[2] A. Kanakia et al., "Mitigating biases in cord-19 for analyzing COVID-19 literature," Frontiers in Research Metrics and Analytics, vol. 5, Nov. 2020. doi:10.3389/frma.2020.596624.

[3] H. Mejia, C. Chipantiza, J. Llumiquinga, I. R. Amaro, and R. Fonseca-Delgado, "Covid-19 fake news detection using joint doc2vec and text features with PCA," Communications in Computer and Information Science, pp. 316–330, 2022. doi:10.1007/978-3-031-20319-0_24.

[4] V. K. Singh, N. Tiwari, and S. Garg, "Document clustering using K-means, heuristic K-means and fuzzy c-means," 2011 International Conference on Computational Intelligence and Communication Networks, Oct. 2011. doi:10.1109/cicn.2011.62.

[5] H. Palangi et al., "Deep sentence embedding using long short-term memory networks: Analysis and application to Information Retrieval," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 24, no. 4, pp. 694–707, Apr. 2016. doi:10.1109/taslp.2016.2520371.

[6] K. D. Onal et al., "Neural information retrieval: At the end of the early years," Information Retrieval Journal, vol. 21, no. 2–3, pp. 111–182, Nov. 2017. doi:10.1007/s10791-017-9321-y

[7] B. Behera, G. Kumaravelan, and P. Kumar B., "Performance evaluation of deep learning algorithms in biomedical document classification," 2019 11th International Conference on Advanced Computing (ICoAC), Dec. 2019. doi:10.1109/icoac48765.2019.246843.

[8] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for Image Classification," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2012. doi:10.1109/cvpr.2012.6248110.

[9] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading text in the wild with Convolutional Neural Networks," International Journal of Computer Vision, vol. 116, no. 1, pp. 1–20, May 2015. doi:10.1007/s11263-015-0823-z.

[10] I. Sutskever and G. Hinton, "Temporal-kernel recurrent neural networks," Neural Networks, vol. 23, no. 2, pp. 239–243, Mar. 2010. doi:10.1016/j.neunet.2009.10.009.

[11] K. Kowsari, M. Heidarysafa, D. E. Brown, K. J. Meimandi, and L. E. Barnes, "RMDL," Proceedings of the 2nd International Conference on Information System and Data Mining, Apr. 2018. doi:10.1145/3206098.3206111.

[12] S. Robertson, "Understanding inverse document frequency: On theoretical arguments for IDF," Journal of Documentation, vol. 60, no. 5, pp. 503–520, Oct. 2004. doi:10.1108/00220410410.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:24 UTC from IEEE Xplore. Restrictions apply.

Language Resources and Evaluation
https://doi.org/10.1007/s10579-024-09753-9

ORIGINAL PAPER

DoSLex: automatic generation of all domain semantically rich sentiment lexicon

Minni Jain[1] · **Rajni Jindal**[1] · **Amita Jain**[2]

Accepted: 5 June 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

For sentiment analysis, lexicons are among the important resources. Existing sentiment lexicons have a generic polarity for each word. In fact, many words have different polarities when they are used in different domain. For the first time, in this work automation of a domain-specific sentiment lexicon named "*DoSLex*" has been proposed. In *DoSLex*, all the words are represented in a circle where the centre stands for the domain, and the *x* and *y* axis for the strength and the orientation of the sentiment, respectively. In the circle, the radius is the contextual similarity between the domain and term calculated using MuRIL embeddings, and the angle is the prior sentiment score taken from various knowledge bases. The proposed approach is language-independent and can be applied to any domain. The extensive experiments were conducted on three low-resource languages: Hindi, Tamil, and Bangla. The experimental studies discuss the performance of the combinations of different word embeddings (FastText, M-Bert and MuRIL) with several sources of prior sentiment knowledge bases on various domains. The performance of *DoSLex* has also been compared with three sentiment lexicons, and the results demonstrating a significant improvement in sentiment analysis.

Keywords Indian languages · Word embedding · WordNet · Lexicon-based methods · Multi-lingual BERT

✉ Amita Jain
amita.jain@nsut.ac.in

Minni Jain
minnijain@dtu.ac.in

Rajni Jindal
rajnijindal@dce.ac.in

[1] Delhi Technological University, New Delhi, Delhi, India

[2] Netaji Subhas University of Technology, New Delhi, Delhi, India

Published online: 18 July 2024

1 Introduction

India is a multi-lingual country and Indians feel comfortable to express their feelings in their native/local language. The number of Indian language internet users across India has reached around 500 million in 2021.[1] This implies that Indians prefer their local languages over English to read, write and converse with each other as they feel more connected and heard. With the ease of internet access, Indian languages such as Hindi, Tamil, Telugu, Bengali, Kannada, and Marathi (to name just a few of the country's 22 official languages) have entered the digital space in a big way. Now it is possible for Indian users to create and consume content in their local language and there is a remarkable rise in content in local languages. Many multi-national companies are also investing in research and resource development for local languages (AL-Husan et al., 2021). The Indian government is also promoting the study and work on regional languages. Despite of the available large amount of content and encouragement, there is a lack of resources in Indian languages (Mohanty, 2008). Developing resources for local/regional languages will usher in an era where there will be an economic boost to the Indian economy. Sentiment analysis can turn out to be beneficial to companies who want to improve the quality of their product or service. Analyzing opinions in the native language can be useful for the Indian government in determining whether the public is satisfied with new policies. Applications of sentiment analysis of social media content can further be seen in entertainment, educational fields, stock market, disaster alerts, management etc.. Sentiment analysis is a natural language processing (NLP) technique that involves determining and extracting sentiments expressed in a piece of text. The primary goal is to understand the subjective information conveyed in the text and categorize it as positive and negative (Joshi et al., 2010).

In the relevant literature there are two widely used approaches for sentiment analysis (a) Lexicon-based and (b) Machine learning based. Machine learning based methods require substantial training data that is representative of the specific domain in which they are applied. They might not perform as well in new or specialized domains without retraining. Whereas Lexicon-based methods can be more domain-independent since they use predefined lists of words with associated sentiments. (Taboada et al., 2011). The proposed work emphasis creating lexicons for sentiment analysis within the framework of lexicon-based methods.

A sentiment lexicon is an important resource composed of sentiment words with polarity and sentiment score (sentimental strength) (Cambria et al., 2015; Esuli & Sebastiani, 2007; Thelwall, 2017). Instances for sentiment lexicons are SentiwordNet (Esuli & Sebastiani, 2007), SenticNet (Cambria et al., 2015), Sentistrength (Thelwall, 2017) etc. HindiSentiWordNet (Joshi et al., 2010) is a sentiment lexicon for the Hindi language. The main notion behind this is that the sentiment of a term is not static, as in traditional lexicon-based approaches, but rather depends on the context or domain in which the term is used, i.e., its contextual semantics. Table 1. illustrates five

[1] https://www.statista.com/statistics/718420/internet-user-base-by-language-india/.

Table 1 Sample sentences having words with different polarity in different domains (with transliteration and English translation of each sentence)

	Domain	Sentence	Word with different polarity	Polarity
1	Movie	फिल्म अनावश्यक रूप से लंबी है *philma anāvashyaka rūpa se laṁbī hai* (The film is unnecessarily long)	लंबा *laṁba* (long)	Negative
	Product (Shampoo)	यह शैम्पू लगाने से बाल लंबे हो जाते है *yaha shaimpū lagāne se bāla laṁbe ho jāte hai* (Applying this shampoo makes the hair long.)		Positive
2	Kitchen(Utensils)	इस बर्तन मे तांबे की मिलावट है *isa bartana me tāṁbe kī milāvaṭa hai* (There is a mixture of copper in this vessel)	तांबे *tāṁbe* (copper)	Neutral
	Jewelry	इस हार मे तांबे की मिलावट है *isa hāra me tāṁbe kī milāvaṭa hai* (There is copper adulteration in this necklace)		Negative
3	Restaurant	रेस्टोरेंट का खाना **ठंडा** था *reṣtoreṁṭa kā khānā ṭhaṁḍā thā* (The food at the restaurant was cold.)	ठंडा *ṭhaṁḍā* (cold)	Negative
	Product (Air Conditioner)	एर कंडीशनर अच्छे से **ठंडा** कर रहा है *era kaṁḍīshanara achchhe se ṭhaṁḍā kara rahā hai* (The air conditioner is cooling well.)		Positive
4	Electronics	फोन **आसानी से** चार्ज हो रहा है *phona āsānī se chārja ho rahā hai* (Phone is charging easily)	आसानी से *āsānī se* (easily)	Positive
	Movie	फिल्म मे आगे क्या होगा **आसानी से** अनुमान लगा**ई** जा सकता था *philma me āge kyā hogā āsānī se anumāna lagāī jā sakatā thā* (What will happen next in the film could have been easily predicted.)		Negative
5	Toy	दोनो कंपनी की कारे **एक जेसी** थी *dono kaṁpanī kī kāre eka jesī thī* (The cars of both the companies were the same.)	एक जेसी *eka jesī* (same)	Neutral

Table 1 (continued)

Domain	Sentence	Word with different polarity	Polarity
Vehicle	दोनो कंपनी की कारे **एक जेसी** थी *dono kaṅpanī kī kāre eka jesī thī* (The cars of both the companies were the same.)		Negative

instances where the polarity of words changes according to the domain. First word is "लंबा" (*laṃbā, length*) which has negative polarity in the movie domain but positive polarity in the domain of product, specifically shampoo or hair treatment products. Similarly, the word "तांबे" (*tāṃbe*, *copper*) has neutral polarity in the case of utensils; on the other hand, it has a strongly negative polarity in the domain of jewellery. From Table 1, it is clearly observable that assigning static polarity to words can degrade the performance of lexicon-based sentiment analysis methods.

In this work, we propose a method that creates a domain-specific lexicon for sentiment analysis. In the proposed approach, the polarity of words is updated by adding contextual semantics of words with the domain. In the proposed *DoSLex*, the lexicon is represented in a geometric 2D circle to identify words' polarity and their sentiment score. The circle's centre is the domain name, and all the terms contextually associated with the domain are plotted in different circle quadrants. Words are classified into four categories, namely positive (words in the first quadrant),

very positive (words in the second quadrant), negative (words in the third quadrant), and very negative (words in the fourth quadrant). The radius in the circle is calculated using FastText, MBert and MuRIL embeddings, and the angle is taken from various knowledge bases. To apply and evaluate proposed *DoSLex*, we also proposed a lexicon-based sentiment analysis method.

The main contributions of this article are listed below:

1. In this research work for efficient sentiment analysis, a semantically rich domain-specific lexicon "*DoSLex*" is proposed.
2. This is the first language independent work for sentiment analysis lexicon generation.
3. The proposed lexicon is evaluated for various domains viz. movie, politics, tourism, books, and news and satisfactory results are obtained with 0.837 F-measure. The proposed approach can be applied to any domain to generate domain-specific lexicon.
4. Unlike state of art sentiment lexicons, the proposed lexicon has been made semantically rich using FastText, M-Bert and MuRIL embeddings

The rest of the article organizes as follows. Section 2 describes the existing state of the art for lexicon generation and sentiment classification methods. The proposed lexicon generation and sentiment classification method is presented in Sect. 3. In Sect. 4, this research shows the experimental results and analysis of the proposed approach. Lastly, Sect. 5 concludes the article with the future direction of lexicon generation research.

2 Related work

Today, after the introduction of UTF-8 encoding style, Hindi content on web is increasing very quickly. Now, it has become easy to type in Hindi, with the help of new keyboards and transliteration resources such as Quillpad (VB et al., 2010)

and Google transliteration. Although there is a lot of content available in the Hindi language, yet the amount of research done in the field of sentiment analysis for Hindi is low. This section discusses the existing literature for lexicons available for both generic or domain-specific sentiment analysis developed for the Hindi language. In the literature, there are different methods for lexicon creation such as manual construction, dictionary-based or corpus- based approaches (Taboada et al., 2011). Manual construction involves human experts creating a sentiment lexicon, who assign sentiment labels (positive, negative, neutral) to words or phrases based on their understanding of the language and context. This method is generally less adopted by researchers because it is time-consuming and it's tough to get sufficient experts (Das & Bandyopadhyay, 2010). Dictionary-based methods utilize a pre-existing small set of seed words with assigned sentiment scores. This list is then extended iteratively by using different approaches (Garg, 2020).

Corpus-based methods involve analyzing large collections of texts (corpora) to identify words and their sentiment scores based on the surrounding context. Machine learning techniques may be employed to extract sentiment information (Medhat et al., 2014). Lexicons are required to contain this type of information to get the polarity of words. Hindi has the Hindi Senti-WordNet (HSWN) (Das & Bandyopadhyay, 2010). HSWN was created manually by IIT Bombay and this resource was developed using English SentiWordNet and linking English-Hindi WordNet. Hindi SentiWordNet consists of synsets that are labelled with positive, negative and neutral sentiment scores attached to their senses. HindiSentiWordNet is a base for many other advances in Hindi sentiment analysis. Next authors automatically and semi-automatically generated a sentiment lexicon using SentiWordNet (Joshi et al., 2010). This lexicon was created for Hindi, Bangla and Telugu languages. A lexicon for adjectives and adverbs was developed by Bakliwal et al. (2012) using Hindi WordNet and they named it HSL (Hindi subjectivity lexicon). Authors used Breadth-first search (BFS) and Hindi WordNet to build HSL. A method to increase the coverage of Hindi Senti WordNet by adding more subjective words was proposed by Mittal et al. (2013). This method also handled negation during the sentiment classification task. The experimental dataset is taken from domain movies only. Next, Arora et al. constructed a sentiment lexicon using the graph theory approach (Arora et al., 2012). A graph is created using WordNet, where nodes are the words and edges are the relations, especially synonymy and antonymy between words taken from WordNet (Arora et al., 2012). This work utilized a pre- annotated small seed list in the beginning. Next Sharma et al. (2014) proposed a bootstrap approach to create the Hindi SentiLexicon that covers Nouns, Adjectives, verbs and adverbs. They developed a multi-module, rule-based, sentiment analysis system. To construct the Hindi SentiLexicon, one positive and one negative seed word are selected. For each keyword from the list, all senses are extracted, and if any of these words have an opposite polarity, the sense is discarded; otherwise, it is added to the lexicon with its polarity. A method by Jha et al. (2016) implemented both lexicon-based classification and machine learning techniques on a Hindi movie reviews dataset. Authors extracted only adjectives and used them as

polarity words for classification. They also used negation handling to improve the accuracy of their methodology. In Mishra et al., (2016) Mishra et al. have built a CSPL (context-specific polarity lexicon) resource for hotel and movie domain in Hindi. They also compared it with standard Hindi SentiWordNet and suggested the inclusion of antonyms, larger datasets, and more. Modi et al. proposed a part-of-speech tagging method using a rule-based approach. For known words, authors used corpus matching and for unknown words, grammar rules were used. This work is for the Hindi language (Modi & Nain, 2016). Jha et al. (2018) developed a sentiment-aware dictionary to handle multiple-domain sentiment classification. The proposed dictionary consists of nouns, adverbs, adjectives and verbs and is aimed to classify amazon product reviews. Hussaini et al. (2018) implemented a score-based sentiment analysis technique with word sense disambiguation handling. They considered book reviews for experimentation. An algorithm for lexicon-based SA is proposed by Yakshi Sharma et al. (2015) and applied to Hindi tweets of "JAI HIND" and "World Cup 2015" topics. A subjectivity lexicon named HSAS (Hindi subjectivity analysis system) was proposed by Jha et al. (2015). It investigates two different strategies for building subjectivity lexicons utilising English language resources, as well as their comparative evaluation in executing the task of subjectivity analysis at the sentence level. The first method makes use of the Opinion-Finder subjectivity lexicon in the English language. The second method starts with a small seed word list in Hindi and develops it to create a subjective lexicon. This second method used PMI (Pointwise mutual information) score and thus expanded 60 seed words to around 4500 polarity words. The PMI (Shapiro et al., 2020) score gives the confidence score of how much is a particular word is associated with a sentiment. Garg et al. (2020) utilized lexicon-based method to perform sentiment analysis of the tweets related to the topic "Mann ki Baat by prime minister Mr. Narendra Modi". Authors used a generic Hindi SentiWordNet lexicon in this approach. Rich Sharma et al. (2014) created a sentiment-oriented system for movie reviews in the Hindi language, they also handled negation in text and achieved 65% accuracy. Authors also compared machine learning and domain-specific-lexicon approach for sentiment analysis (Gupta et al., 2021). Table 2 compares various research works on lexicon-based sentiment analysis and sentiment lexicon generation. The table explains the purpose of the work, the language, the domain used and the accuracy obtained. The accuracy mentioned in fifth column of Table 2 has been taken from the experimentation section of associated research papers.

3 Proposed methodology for generation of *DoSLex* and lexicon-based sentiment analysis

This section introduces the framework for the generation of the domain-specific sentiment lexicon "*DoSLex*". The same steps can be applied to any language for lexicon generation. *DoSLex* aims to represent word sentiment polarity (positive, very positive, negative and very negative) with its associated sentiment score using contextual

Table 2 Comparison table of various lexicon generation and lexicon-based sentiment analysis approaches with Hindi, Bangla and Tamil low resource languages

Reference	Purpose	Main feature	Domain application	Accuracy	Language
Das and Bandyopadhyay (2010a)	Generic Lexicon Generation	Generation of sentiment lexicons using WordNet	Random	Positive-88% Negative-91%	Bangla, Hindi, Telugu
Joshi et al. (2010)	Generic Lexicon Generation	Development of Hindi Senti WordNet	Movie Reviews	60.31%	Hindi
Mittal et al., (2013)	Generic Lexicon Generation	Improvement of existing Senti Hindi WordNet by negation handling and discourse relations	Movie Reviews	80.21%	Hindi
Arora et al. (2012)	Generic Lexicon Generation	WordNet Graph traversal for lexicon generation	Amazon products reviews	74.1%	Hindi
Sharma and Bhattacharyya (2014)	Generic Lexicon Generation and Sentiment Classification	Lexicon through bootstrap approach and multimodal rule based SA system	Movie and Products Reviews	87%	Hindi
Jha et al. (2016)	Sentiment Classification	Opinion word extraction using adjectives	Movie Reviews	Positive-93% Negative-92%	Hindi
Mishra et al. (2016)	Domain Specific Lexicon Generation	Dictionary based context sensitive polarity lexicon for specific domain	Movie and Hotel Reviews	Movie-88% Hotel-77%	Hindi
Modi and Nain (2016)	Part of speech Tagging system	Rule-based approach for POS tagging a corpus	News and stories	85.4%	Hindi
Jha et al. (2018)	Domain Specific Lexicon Generation	Dictionary for multi-domain sentiment analysis	Amazon products reviews	76.2%	Hindi
Hussaini et al. (2018)	Sentiment Classification	Sentiment scoring with word sense disambiguation	Book Reviews	87.4%	Hindi

Table 2 (continued)

Reference	Purpose	Main feature	Domain application	Accuracy	Language
Sharma et al. (2015)	Generic Lexicon Generation and Sentiment Classification	Constructed a subjective lexicon and classify Hindi Tweets	Tweets (World cup 2015 and Jai Hind)	World Cup 2015-81.9% and Jai Hind-73.5%	Hindi
Jha et al. (2015)	Lexicon Generation and Sentiment Classification	Constructed lexicon using English: opinion finder and PMI approach	Hindi and English Dataset	80%	Hindi
Garg (2020)	Sentiment Classification	Dictionary-based approach for SA of tweets of PMs Mann ki Baat	Tweets (PM's Mann ki baat)	84.2%	Hindi
Sharma et al. (2014)	Sentiment Classification	Hindi sentiment Orientation System along with negation handling	Movie Reviews	65%	Hindi
Pravalika et al. (2017)	Sentiment Classification	SA for bi-lingual sentence using lexicon and machine learning approach	Facebook	88%	Hindi-English Code-mixed
Garg and Lobiyal (2020)	Emotion lexicon generation and classification	Hindi EmotionNet is built using graph centrality measures	screenplays, stories, and blogs	85.78%	Hindi
Sivasankar et al. (2021a)	Domain Specific Lexicon Generation	Domain specific lexicon generation using PMI	Amazon Product Reviews	84%	Tamil
Sazzed (2021)	Generic SA Lexicon Generation and profanity detection	Lexicon generation and hate speech detection using PMI and ML classifiers	Drama and News	65%	Bangla

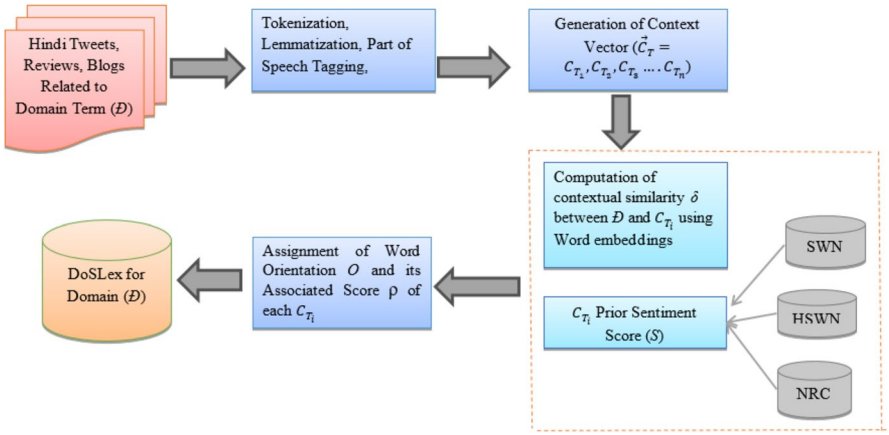


Fig. 1 Overall Framework for Hindi *DoSLex* Generation

semantic analysis and word embeddings. After the *DoSLex* generation, a lexicon-based sentiment analysis was conducted to evaluate the lexicon's performance. This lexicon generation framework considers the six main steps that are shown in Fig. 1. For the better understanding and ease of explanation of the proposed methodology, the illustrations are given for Hindi language.

3.1 Data collection and pre-processing

The proposed approach for lexicon building can be applied to any domain but in this paper to make it feasible, we have selected five Hindi domains, one Tamil and one Bangla namely movie, restaurant, politics, books and tourism, Tamil movie and Bangla news. Sentences related to these domains are extracted from various sources including Twitter,[2] News websites,[3,4,5,6] IIT Bombay site. After extraction, the first pre-processing step is the lemmatization and tag extraction of data according to suitable part of speech. For lemmatization, we have used pyspark from John snow labs[7] and for part of speech tagging NLTK POS tagger is used.

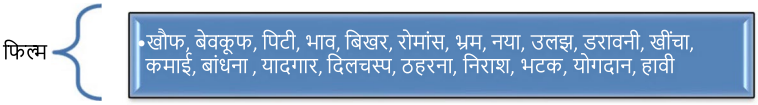
[2] www.twitter.com.
[3] www.Jagran.com.
[4] www.raftaar.in/.
[5] www.amarujala.com/.
[6] www.bbc.com/hindi.
[7] www.johnsnowlabs.com.

Algorithm 1 Domain Specific Lexicon (*DoSLex*) Generation

Input: Dataset related to domain ($Đ$) from various sources for language L
Output: Domain-Specific Lexicon *DoSLex* comprising polarity and sentiment score for each word. (classification of words in four polarity: positive, Very positive, Negative and Very Negative)
Initialize POS^{+} List, $VPOS^{++}$ List, NEG^{-} List, $VNEG^{--}$ List to empty
for each sentence in dataset
Perform Lemmatization and Tokenization
for each token do
Part-of-Speech tagging (POS)
if POS= nouns||adjectives|| adverbs || verbs || Unknown then
Add token in to context vector $\overrightarrow{(C_T} = C_{T_1}, C_{T_2}, C_{T_3} \ldots . C_{T_n})$
for each context term $C_{T_i} \in \vec{C}_T$
Calculate $\delta\,(Đ, C_{T_i})$ // δ is semantic relation between domain term and context term
Calculate $S(C_{T_i})$ // S is prior sentiment of context term
Plot in 2D circle
$A_i = \delta\,(Đ, C_{T_i})$ // radius $0 \leq A_i \leq 1$
$\emptyset_i = S(C_{T_i}) * \pi$ //angle
Coordinates for each context term
$X_i = A_i \sin \emptyset_i$ // polarity score of context term
$Y_i = A_i \cos \emptyset_i$ // polarity of context term
if (X_i, Y_i) lies in first quadrant
Add C_{T_i} in POS^{+} List
Elseif (X_i, Y_i) lies in Second quadrant
Add C_{T_i} in $VPOS^{++}$ List
Elseif (X_i, Y_i) lies in Third quadrant
Add C_{T_i} in $VNeg^{--}$List
Elseif (X_i, Y_i) lies in Fourth quadrant
Add C_{T_i} in Neg^{-} List
Return *DoSLex*, POS^{+} List, $VPOS^{++}$ List, NEG^{-} List, $VNEG^{--}$ List

3.2 Generation of context vector ($\vec{C}_T$)

A context vector is denoted as $(\vec{C}_T = C_{T_1}, C_{T_2}, C_{T_3} \ldots C_{T_n})$ where, $C_{T_i} \in \vec{C}_T$ and C_{T_i} denotes a context term. Context terms are words, related to the domain $Đ$ whose polarity we are going to assign using proposed approach. This work considers all terms that occur in the same context to generate a lexicon for each domain $Đ$. Let's say if domain $Đ$ is "फिल्म" (*Filma, Movie*), then all the sentences related to the domain are extracted from different platforms (in the previous step). This implies that acontext vector comprises contextual terms of domain. For domain "फिल्म" (*Filma, Movie*) a sample context vector is shown in following figure.



3.3 Computation of semantic relation (δ) between $Đ$ and C_{T_i} using word embeddings

To compute semantic relation (**δ**) between domain $Đ$ and *CTi*, we have utilized the word embedding method. A word embedding is a term used for the vector representation of words and phrases with similar words and phrases having similar word representation in a predefined vector space (Gaikwad & Haribhakta, 2020). FastText is one of the effective word embeddings for the Hindi language (Kumar et al., 2020). FastText is an open-source, free, lightweight library that allows users to learn text representations and text classifiers. FastText (essentially an extension of the word2vec model) treats each word as if it is composed of character n-grams. That means that the vector for a word is made up of the sum of its character n-grams. For the work discussed here, we initially applied FastText embeddings. However, since the results were not satisfying, we shifted to M-Bert (Multilingual BERT) and then finally to the latest embedding method for Indian languages MuRIL (Multilingual Representations for Indian Languages) (For results, please refer to Sect. 4.2). Multilingual Bidirectional Encoder Representations from Transformers (M-Bert) is a pre-trained language model developed using the 104-language Wikipedia corpus (Pathak et al., 2021). The main accomplishment of this model is its ability to transfer across languages with different vocabularies, such as English and Devanagari Hindi, even though it was trained on distinct 104 monolingual corpora beforehand. Additionally, empirical evidence suggests that M-Bert is able to detect semantic relationships between words. Another model built on Bert that is pre-trained in Indian languages is MuRIL (Khanuja et al., 2103). It is a recently released BERT model that has been pre-trained on 17 Indian languages, including Hindi and English, as well as their transliterated equivalents. It is a modified version of M-Bert for Indian languages with limited resources because pairs of translated and transliterated texts have been used for its training.

3.4 Prior generic sentiment score (*S*)

In this step, each context word $C_{T_i} \in \overrightarrow{C}_T$ is assigned an initial score from external sentiment lexicons. Here sentiment scores for each Hindi context word are taken from the Hindi SentiWordNet lexicon (Joshi et al., 2010), NRC-Emotion lexicon (Bose et al., 2019; Mohammad & Turney, 2013) and English SentiWordNet. For a maximum coverage of words, more than one lexicon has been considered. The words that were not listed the Hindi Senti WordNet and NRC emotion lexicon were translated to English and checked in the English Senti WordNet. For Bangla and Tamil languages, Bengali SentiWordNet and Tamil SentiWordNet are utilized.

3.5 Computation of word orientation *O* and its associated sentiment score ρ of each C_{T_i}

Each domain $Đ$ and its context term $C_{T_i} \in \overrightarrow{C}_T$ are represented in a geometric 2D circle. This circular representation is beneficial as it allows us to use trigonometric

identities to estimate the sentiment orientation and strength of the various context terms. *DoSLex* is based on contextual semantic relations within a domain *Đ*. In this geometric circle, domain *Đ* is the centre of the circle and the terms in domain *Đ* are points found around it. The position of a context term C_{T_i} is defined jointly using word embeddings and prior sentiment scores. Mathematically, a 2D circle polar coordinate system is represented with the following equation:

$$\mathrm{A}^2 - 2\mathrm{A}\mathrm{A}_0 \cos(\emptyset - \theta) + \mathrm{A}_0^2 = \mathrm{R}^2 \tag{1}$$

where *R* denotes the radius of the circle, the polar coordinate of a domain term *Đ* that situated in centre of the circle as (A_0, θ) and the polar coordinate of a context term C_{T_i} as (A, $\varnothing$). This study considers context wordC_{T_i}, radius A_i, angle $\varnothing_i$ to build *DoSLex* of *Đ*, Prior of sentiment score *S* and word embedding semantic relationship score δ (Arora et al., 2012) of word C_{T_i}. The Eqs. (2) and (3) help build the 2D circle for *DoSLex*.

$$\mathrm{A}_i = \delta(Đ, C_{T_i}) \tag{2}$$

$$\emptyset_i = S(C_{T_i}) * \pi \tag{3}$$

The radius of all terms in a 2D circle is considered a scale between 0 and 1, and the circle's radius R is 1. All angle values are computed in radians.

3.6 Generation of *DoSLex*

DoSLex uses the Cartesian coordinate system from the polar coordinate system by applying the sineand cosine trigonometric functions to compute the coordinates of term C_{T_i}.

$$X_\mathrm{i} = \mathrm{A}_i \sin\emptyset_i$$

$$\mathrm{Y}_\mathrm{i} = \mathrm{A}_i \cos\emptyset_i$$

The Y-axis of the Cartesian coordinate system denotes sentiment orientation (positive, very positive, negative and very negative) of term C_{T_i} and the X-axis denotes the degree (score) of the C_{T_i} term in the context of the domain. A positive Y value stands for positive sentiments and negative Y value for negative sentiments. A small X value denotes strong sentiment as shown in Fig. 2.

Figure 3 represents a sample 2D circle representation of *DoSLex* of domain movie, tourism, books and news for the Hindi language. Translation and Transliterations of each illustration are provided in Appendix 1. In this figure, terms present in the first and second quadrant are positive terms with respect to domain and terms in the third and fourth quadrant are negative terms. This domain-specific lexicon is a very significant resource for research in sentiment analysis. It provides context words strongly related to the domain with polarity and ignores context words that

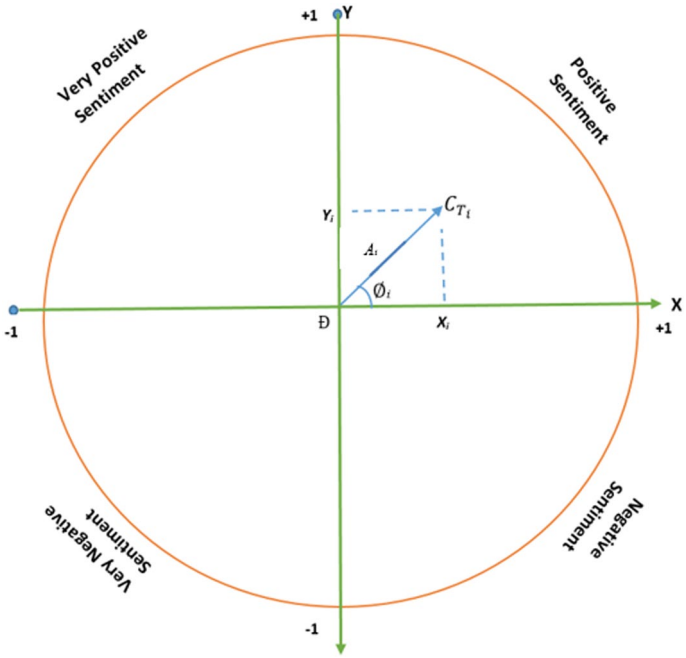


Fig. 2. 2D circle representation for *DoSLex* with Four Polarity quadrants

contribute low polarity scores in the particular domain. In Fig. 3, the 2D circles for "किताब" (Kitāba, book) and "समाचार" (samāchāra, news) share the term "शांति" (Shāanti, Peace) but with varying polarities. In the 2D circle for "किताब" (Kitāba, book), "शांति" (Shāanti, Peace) has a positive polarity, while in the circle for "समाचार" (samāchāra, news), it has a very positive polarity. Number of homonyms or the terms having different polarities in dataset is approx. three hundred forty-five terms.

3.7 Lexicon-based sentiment analysis using *DoSLex*

In this section, we discussed a methodology for lexicon-based sentiment analysis using *DoSLex* for Hindi language. Figure 4. presents the overall workflow diagram for the sentiment analysis task. Regarding the computation of sentiment score, the first step is to pre-process the extracted text T. Input text can be taken from any source such as reviews, tweets etc. Pre-processing includes removal of URLs, abbreviations, punctuations and symbols. Next step is the identification of the domain using topic modeling LDA (Latent Dirichlet Algorithm) (Jelodar et al., 2019). LDA is a generative probabilistic model used for topic modeling (Jelodar et al., 2019). LDA is an unsupervised learning algorithm, it doesn't require labeled training data and it can be applied to a variety of textual data, such as articles, books, or social media posts (Jelodar et al., 2019). The identified domain is then searched in

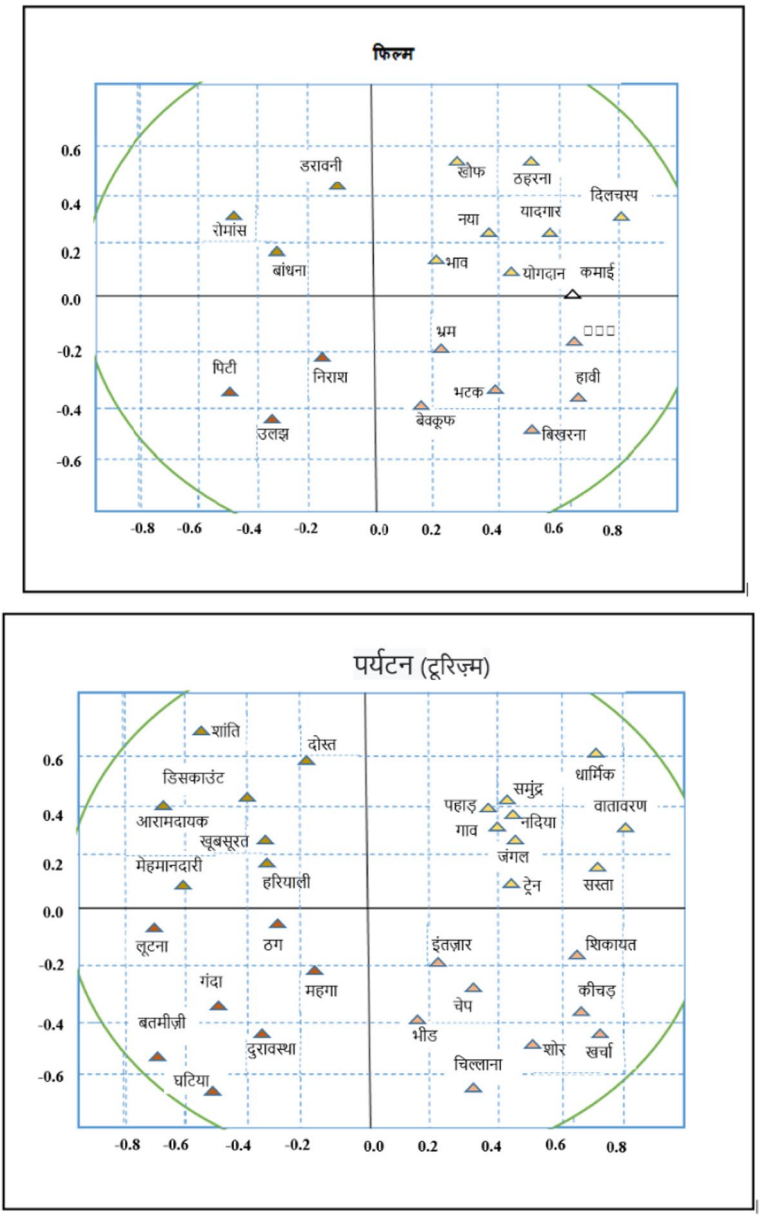


Fig. 3 *DoSLex* 2D circle representation for domains फिल्म (movie), पर्यटन (Tourism), किताब (book) and समाचार (news) in Hindi

to proposed list of lexicons (for instance in this proposed we have created lexicons for movies, tourism, news, politics, books domain). In case of non-availability of domain, system checks it's synonym in *DoSLex*. And if neither a synonymy is available, then algorithm 1 is applied to generate the required lexicon. After getting the

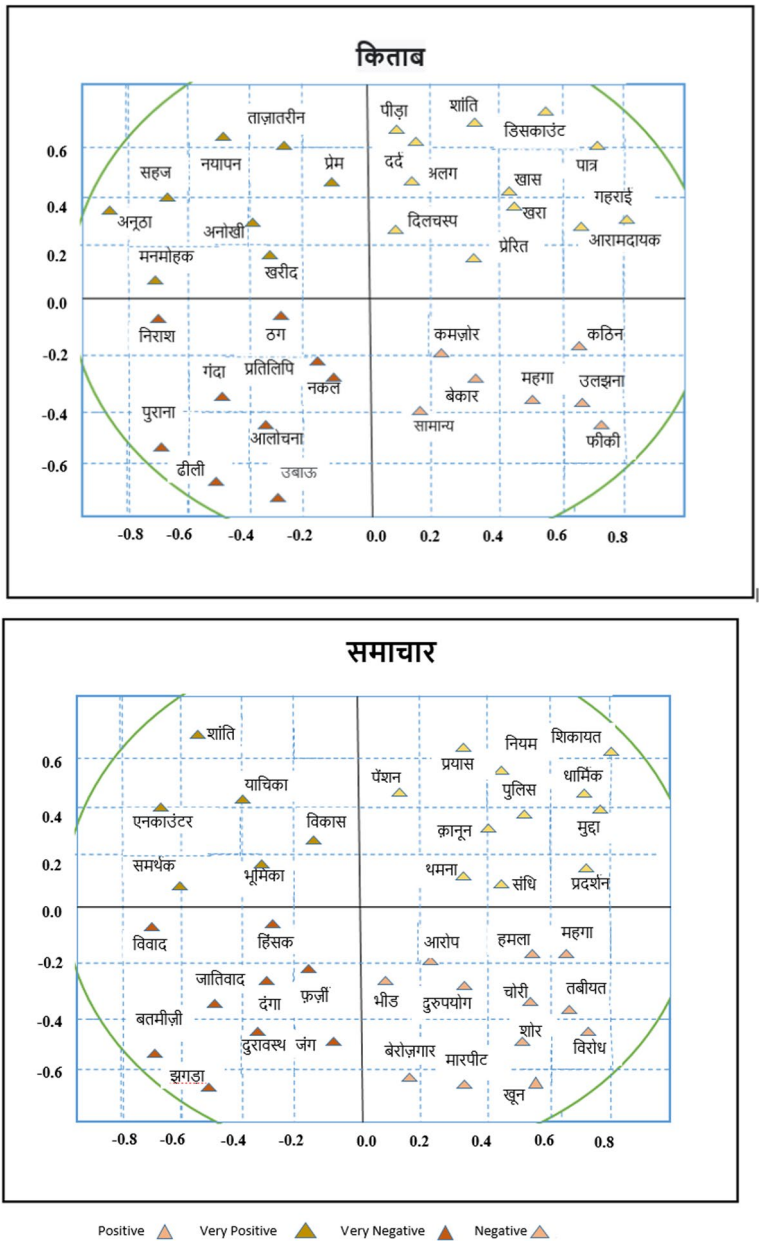


Fig. 3 (continued)

relevant domain specific lexicon,the text is split to sentences ($T = S_1, S_2, S_3, \ldots S_p$) where T is the text and $S_i \epsilon T$ is the sentence of text. Each sentence is then tokenized and each token tagged according to appropriate part of speech. For each noun,

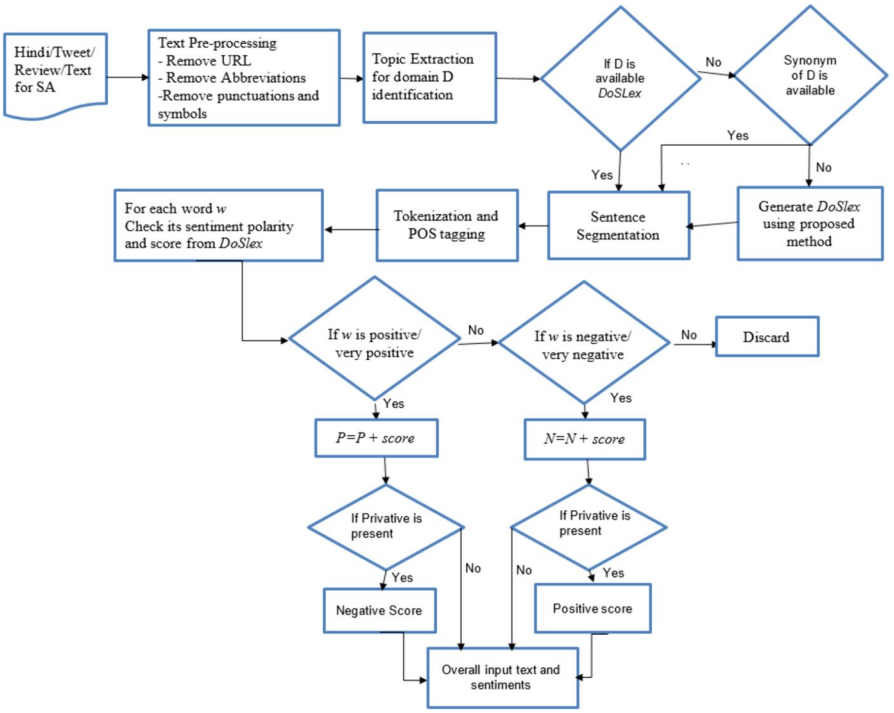


Fig. 4 Overall flowchart for lexicon-based sentiment analyses approach using *DoSLex*

adverb, adjective and verbs tag we check the sentiment polarity and strength in *DoSLex*. So now, there are two categories of sentiment score: one is positive and other is negative. In case negation/privative is present in the sentence, its polarity will be reversed. This step normalizes the impact of negations in a sentence. For example, in the sentence "I am not satisfied with the product," the presence of "not" (negation) alters the polarity of the sentence from positive to negative. In last step all the sentiment scores are aggregated and an overall positive score/negative score of sentence is obtained.

4 Experimental results and discussion

This section the efficiency of the proposed method is tested with several experiments. To this end, evaluation matrices, datasets description and experiment details with results are presented and discussed. This research utilizes Quad core Intel Core i7 Skylake system with16GB of RAM. The model is implemented in Keras and TensorFlow (TensorFlow 2018).

4.1 Dataset description and evaluation matrices

Obtaining annotated datasets for low-resource languages is a challenging task. Although the proposed approach can be applied to any language, yet for ease and understandability, three Indian languages Hindi, Tamil and Bangla are used. This section discusses the dataset for Hindi, Bangla and Tamil languages. Hindi Dataset by IIT Bombay: Annotated dataset for domain tourism and movie in Hindi language is taken from IIT B site. The movie domain consists of 252 positive and negative labelled reviews (Joshi et al., 2010). The tourism domain has 200 positive and negative reviews (Balamurali et al., 2012).

Twitter Dataset: The dataset for evolution of proposed methodology has been fetched from twitter using the tweepy API. The 2467 tweets have been extracted for "राजनीति" (politics) domain from twitter by selecting language Hindi in the search filter. Out of these tweets, we removed tweets with English words written in Hindi, non-Hindi tweets and tweets without subjectivity. Tweets without subjectivity in annotation are tweets presenting factual information without expressing personal feelings or opinions. They are typically neutral and do not convey sentiment.

Example: "The smartphone has a 6-inch display and a 12-megapixel camera." The proposed work is handling positive/very positive, negative/very negative sentiments After elimination 1534 subjective tweets were left. These tweets were given to three annotators for manual labelling. The annotators were graduate students and native speakers of Hindi language. They have been asked to label each tweet with the lables "positive" or "negative" according to their understanding. Now, different annotators may have different perceptions of meaning convey through context.

Cohen's Kappa was used to measure inter-annotator agreement. This is a statistical measure used to assess the inter-annotator agreement or reliability between two or more annotators when labelling categorical items. It accounts for the agreement occurring by chance. This measure is widely used in various fields, including natural language processing (NLP), to measure agreement on categorical annotations.

Disagreement Handling: In case of disagreement, the examples at stake were excluded from the corpus.

We also used Gwet's AC1 which is designed to address limitations of Cohen's Kappa.

Finally, a dataset of 1534 tweets belonging to domain "politics" were extracted; of them 654 tweets were assigned the label "positive" and 880 the label "negative".

Amazon Product review dataset: It is a benchmark dataset which is widely used for sentiment analysis task. All the reviews are collected from amazon.com. There are 1000 positive and 1000 negative reviews for each product. We collected for domain "किताब" (*Kitabh*, books) and translated them using Google translator.

News Comment Dataset for Bangla: This dataset is collected from two popular Bangla newspapers, BBC Bangla and Prothom Alo. It contains 5205 positive comments and 5600 negative tagged comments (Taher et al., 2018).

Tamil Movie Reviews: This dataset consists of 200 positive and 200 negative comments reviews from Movie dataset (Film: Baahubali) (Sivasankar et al., 2021b). Table 3 shows a detailed description of annotated dataset of all domains.

Table 3 Annotated Dataset Description

Language	Domain	Positive Annotated dataset	Negative Annotated dataset
Hindi	Books	1000 Reviews	1000 Reviews
	Movies	127 Reviews	125 Reviews
	Tourism	100 Reviews	100 Reviews
	politics	654 Tweets	880 Tweets
Bangla	News	5205 Comments	5600 Comments
Tamil	Movie	200 Reviews	200 Reviews

The evaluation matrices considered in this research are precision (P_R), recall (R_E) and F-measure score (F_M). These matrices will assess the performance of the proposed Domain-Specific lexicon. The P_R is the fraction of correctly classified sentences of a sentiment (Positive or negative) among number of sentences classified by the relevant sentiment. The R_E is fraction of correctly classified sentence count among the total number of sentences of that sentiment. The R_E and P_R helps to understand the degree of relevance. The F score is the harmonic mean of the P_R and R_E.

4.2 Results analysis of the proposed *DoSLex* on various datasets

This section describes in detail the experimental results that will help to comprehend how the proposed Domain-Specific lexicon performed well in sentiment classification tasks. For evaluation, the lexicon-based approach discussed in Sect. 3.7 has been applied to different domains. We conducted experiments using HSWN, followed by NRC, SWN, and *DosLex*, as outlined in Table 4. The performance of *DosLex* on the dataset is notably superior because these lexicons offer generic polarity for each term without considering the specific domain in which the term appears. For example, the word 'long' has a generic positive polarity in SentiWordNet, but *DosLex*, provides polarity scores based on the domain context (negative in the case of a movie like 'long movie' and positive in the case of a product like 'long hair' shampoo). In many instances, HSWN, NRC, and SWN yield inaccurate results. We compute the performance of *DoSLex* with different sources of prior sentiment knowledge bases. First is *DoSLex* with HSWN only. Second is *DoSLex* with HSWN + NRC and the last one is *DoSLex* with HSWN + NRC + ESWN. Table 5 presents the evaluation of all the cases. It can be noticed from the table that *DoSLex* with HSWN achieves maximum F-measure of only 0.306 in domain of Hindi movies. The performance of *DoSLex* with HSWN is very low due to narrow coverage of polarity words in Hindi SentiWordNet related to the domain HSWN has very limited polarity words in the lexicon. Out of this polarity some words are wrongly labelled for example word "हिंसा" (*hinsa,* Violence) is marked with neutral polarity instead of negative polarity. However, *DoSLex* with HSWN + NRC gives maximum F-measure of 0.542 due to increase in polarity word coverage and reconsideration of neutral words. Out of these three cases, *DoSLex* with HSWN + NRC + ESWN (FastText)

Table 4 Performance evaluation results of proposed *DoSLex* with different existing sentiment bases

Dataset	Domain	HSWN			NRC			SWN			Proposed lexicon		
		P_R	R_E	F_M	P_R	R_E	F_M	P_R	R_E	F_M	P_R	R_E	F_M
IIT B	Movie	0.160	0.181	0.169	0.192	0.218	0.204	0.230	0.268	0.248	0.829	0.836	0.832
	Tourisms	0.158	0.140	0.149	0.168	0.174	0.171	0.211	0.194	0.202	0.833	0.843	0.837
Amazon	Books	0.176	0.174	0.175	0.179	0.163	0.171	0.192	0.20	0.196	0.824	0.835	0.829
Twitter	Politics	0.147	0.125	0.135	0.132	0.127	0.129	0.262	0.234	0.247	0.842	0.831	0.836

Table 5 Performance evaluation results of proposed *DoSLex* with different prior sentiment bases and FastText

Dataset	Domain	*DoSLex* with HSWN (FastText)			*DoSLex* with HSWN + NRC (FastText)			*DoSLex* with HSWN + NRC + SWN (FastText)		
		P_R	R_E	F_M	P_R	R_E	F_M	P_R	R_E	F_M
IIT B	Movie	0.314	0.299	0.306	0.563	0.523	0.542	0.689	0.679	0.683
	Tourisms	0.271	0.253	0.261	0.512	0.553	0.531	0.623	0.653	0.637
Amazon	Books	0.262	0.234	0.247	0.534	0.434	0.478	0.644	0.634	0.638
Twitter	Politics	0.212	0.22	0.215	0.502	0.450	0.474	0.612	0.6	0.605

Table 6 Performance evaluation results of proposed *DoSLex* with different prior sentiment bases and M-Bert

Dataset	Domain	*DoSLex* with HSWN (M-Bert)			*DoSLex* with HSWN + NRC (M-Bert)			*DoSLex* with HSWN + NRC + SWN (M-Bert)		
		P_R	R_E	F_M	P_R	R_E	F_M	P_R	R_E	F_M
IIT B	Movie	0.332	0.326	0.328	0.643	0.646	0.644	0.734	0.767	0.750
	Tourisms	0.317	0.335	0.325	0.613	0.672	0.641	0.712	0.734	0.722
Amazon	Books	0.319	0.345	0.331	0.667	0.656	0.661	0.745	0.756	0.750
Twitter	Politics	0.321	0.322	0.321	0.623	0.634	0.628	0.745	0.712	0.728

Table 7 Performance evaluation results of proposed *DoSLex* with different prior sentiment bases and MuRIL

Dataset	Domain	*DoSLex* with HSWN (MuRIL)			*DoSLex* with HSWN + NRC (MuRIL)			*DoSLex* with HSWN + NRC + SWN (MuRIL)		
		P_R	R_E	F_M	P_R	R_E	F_M	P_R	R_E	F_M
IIT B	Movie	0.399	0.419	0.408	0.773	0.753	0.762	0.829	0.836	0.832
	Tourisms	0.402	0.423	0.412	0.762	0.775	0.768	0.833	0.843	0.837
Amazon	Books	0.372	0.364	0.367	0.754	0.774	0.763	0.824	0.835	0.829
Twitter	Politics	0.362	0.352	0.356	0.72	0.761	0.739	0.842	0.831	0.836

achieves a better F-measure of maximum 0.683 in domain movie. This performance depends on prior sentiment values and can be improved further with a more accurate and large size of sentiment lexicon. The proposed approach is also reliant on the performance of word embeddings including FastText. M-Bert and MuRIL. Tables 5, 6 and 7 provides the results of all combinations with three word embeddings. Figure 5 represents the f-measure comparison of FastText, M-Bert and MuRIL embeddings on different domain datasets. Figures shows M-Bert performs better than FastText because in FastText each word has fixed representation regardless of context and M-Bert is a contextual based word embedding. MuRIL performs best out of

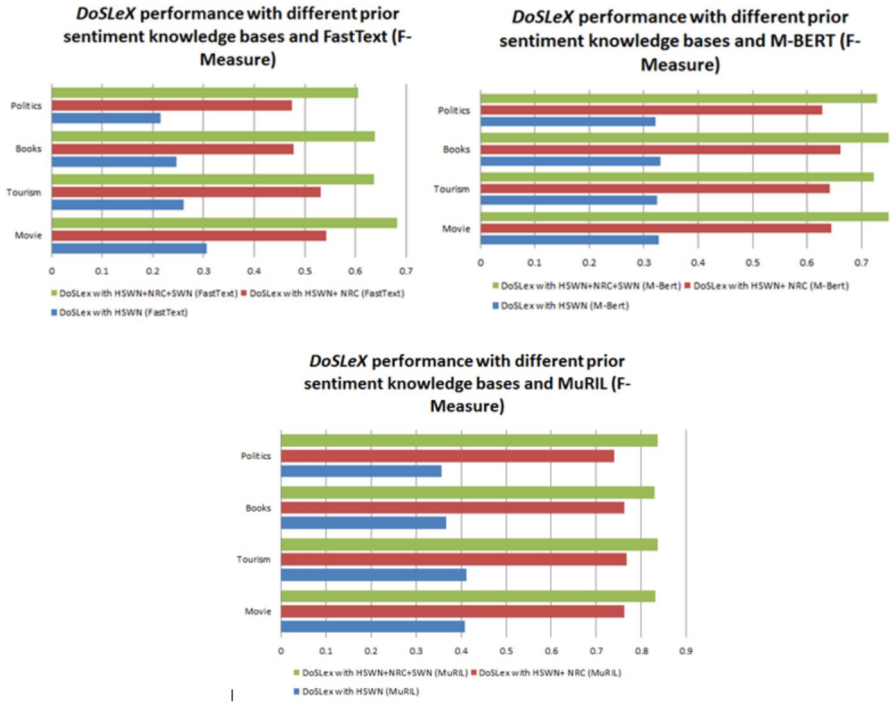


Fig. 5 Graphical representation for the performance of *DoSLex* with different prior sentiment knowledge bases and embeddings (F-measure)

Table 8 Performance evaluation of *DoSLex* for Bangla and Tamil low-resource languages

Domain and Language	Model Combinations	P_R	R_E	F_M
Bangla News	*DoSLex* + BSWN + FastText	0.323	0.337	0.329
	DoSLex + BSWN + M-Bert	0.572	0.563	0.567
	DoSLex + BSWN + MuRIL	0.685	0.677	0.680
Tamil Movie	*DoSLex* + TSWN + FastText	0.381	0.372	0.376
	DoSLex + TSWN + M-Bert	0.591	0.601	0.595
	DoSLex + TSWN + MuRIL	0.721	0.703	0.711

these three embedding techniques. MuRIL has similar training strategy as M-Bert with few modifications Maximum F-measure is 0.837 given by *DoSLex* + HSWN + NRC + SWN with MuRIL embedding. Tables 5, 6 and 7 show the result for Hindi language.

Next experiments are done for Bangla and Tamil languages for news and movie domain. Table 8 shows results with different combinations. For Bangla news domain *DoSLex* + BSWN + MuRIL combination provided best performance with 0.68 F-measure. And for Tamil movie domain *DoSLex* + TSWN + MuRIL gives best

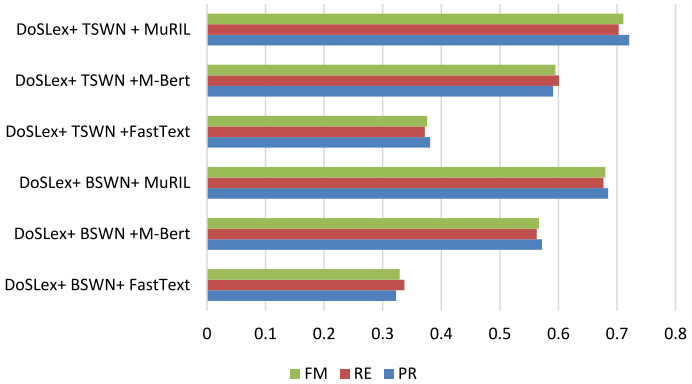


Fig. 6 Graphical representation for performance of DoSLex for Bangla and Tamil language

Table 9 Performance evaluation results of proposed *DoSLex* with Indic_SWN on various datasets

Dataset	Domain	*DoSLex* (Proposed)			Indic_SWN (HSWN, BSWN, TSWN)		
		P_R	R_E	F_M	P_R	R_E	F_M
IIT B	Movie	0.829	0.836	0.832	0.160	0.181	0.169
	Tourisms	0.833	0.843	0.837	0.158	0.140	0.149
Amazon	Books	0.824	0.835	0.829	0.176	0.174	0.175
Twitter	Politics	0.842	0.831	0.836	0.147	0.125	0.135
	Bangla News	0.685	0.677	0.680	0.120	0.131	0.125
	Tamil Movie	0.721	0.703	0.711	0.061	0.134	0.083

results of 0.711 f-measure. BSWN and TSWN are Bangla SentiWordNet and Tamil SentiWordNet respectively. The results are shown in graphs in Fig. 6. The standard/classic Indic_sentiwordNet (SentiWordNets for Indian languages such as Hindi, Bangla, Tamil) are compared with proposed *DoSLex*. The comparison with existing popular generic-lexicons is performed on different domains to show the performance of proposed *DoSLex* over existing lexicons. For experimentation, same sentiment classification method explained in Sect. 3.7 is applied to all datasets. Table 9 and Fig. 7. shows the evaluation results and from the table it can be noticed that proposed *DoSLex* performs much better with maximum F-measure of 0.837, 0.680, 0.711 in Hindi, Bangla and Tamil languages respectively. The major reason of very low performance of generic lexicons is non consideration of topic or domain of the text. For instance, "small" is assigned negative in SentiWordNet but it is positive in many domains such as mobile "small and handy". Moreover, the proposed lexicon is extendable to any number of polarity words depending on prior knowledge and word similarity and *DoSLex* is language independent. *DoSLex* is also fully scalable and adaptable to any domain. Figures 5, 6 and 7 show the graphical representation of whole experimental results. Indic_SWNs, encompassing Hindi WordNet, Bangla WordNet, and Tamil WordNet, cannot be deemed reliable due to a notably low

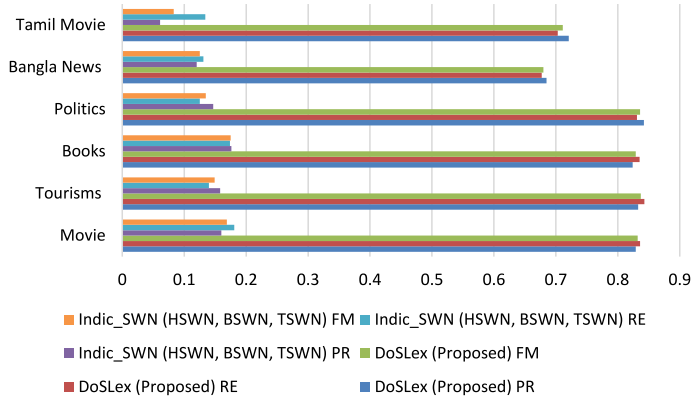


Fig. 7 Graphical representation for comparative results of proposed *DoSLex* with indic_sentiwordnets on various datasets

F-measure, falling below 0.2. However, the proposed lexicon, when amalgamated with these Indic_SWNs, yields satisfactory results with an F-measure of 0.83.

5 Conclusion and future work

In this paper, the authors describe the lexical resource *DoSLex which is* a domain-specific sentiment lexicon. The methodology leverages domain-related corpora, word embeddings and prior sentiment knowledge bases. In this work, it has been found that representing lexicons in the 2D geometric circle is helpful as it allows to use trigonometric formulas to calculate sentiment orientation and strength of words in a given context. The proposed method is applied to three low-resource languages Hindi, Bangla and Tamil and can be applied on any other languages such as Sanskrit, Marathi etc. It can also be extended effortlessly to any number of domains. The experimentation section revealed that the proposed method performs well in different domains (Movie, Books, Politics, Tourism, Tamil movie, Bangla News). The method is validated for three languages and it is found that *DoSLex* performs much better than traditional generic sentiment lexicons. A key advantage of the proposed method is its language independence, making it applicable for generating lexicons in any domain. Following the similar approach, lexicons can also be generated for agriculture, healthcare incorporating depression detection, hate speech detection, pandemic support, pharmacy, disease detection & treatment, and vaccine performance evaluation, which do not exist right now. It would be interesting to see the implementation and performance of *DoSLex* for lexicon generation in these specific areas. In the future, it would be interesting to work on lexicon generation for low-resource code mixed languages such as Hindi-English, Punjabi-English etc. One drawback of the proposed approach is that it doesn't account for the influence of modifiers and adjectives on the polarity of a term. An interesting future research direction may be the investigation of modifiers. adjectives, slang terms, metaphors and sarcasm in the

text to assign word polarity more accurately. Handling word sense disambiguation in languages is another area that can be done in future for better performance.

Appendix 1

Word	Transliteration	Translation	Word	Transliteration	Translation
खौफ	*Khaufa*	Awe	डिसकाउंट	*ḍisakāuanṭa*	Discount
बेवकूफ	bevakūfa	Stupid	आरामदायक	*aarāmadāyaka*	Comfortable
पिटी	*piṭī*	Pity	खूबसूरत	*khūbasūrata*	Beautiful
भाव	*bhāva*	Emotion	मेहमानदारी	*mehamāna dārī*	Hospitality
बिखर	*bikhara*	Shatter	हरियाली	*hariyālī*	Greenery
रोमांस	*romāansa*	Romance	लूटना	*lūṭanā*	Robbery
भ्रम	*bhrama*	Confusion	गंदा	*gandā*	Dirty
नया	*nayā*	New	महगा	*maheangā*	Expensive
उलझ	*ulaza*	Tangled	ठग	*ṭhaga*	Cheating
डरावनी	*ḍarāvanī*	Scary	दुरवस्था	*duravasthā*	Plight
खींचा	*khīanchā*	Brawn	समुंद्र	*samuandra*	Sea
कमाई	*kamāee*	Earn	धार्मिक	*dhārmika*	Religious
बांधना	*bāandhanā*	Binge	वातावरण	*vātāvaraṇa*	Environment
यादगार	*yādagāra*	Memorable	नदिया	*nādiyā*	River
दिलचस्प	*dilachaspa*	Interesting	पहाड़	*pahāḍa*	Mountain
ठहरना	*ṭhaharanā*	Stay	सस्ता	*sastā*	Cheap
निराश	*nirāsha*	Frustrated	ट्रेन	*ṭrena*	Train
भटक	*bhaṭaka*	Wander	जंगल	*jangala*	Jungle
योगदान	*yogadāna*	Contribute	इंतज़ार	*antajaāra*	Waiting
हावी	*hāvī*	Dominate	चेप	*chepa*	Attached
शांति	*Shāanti*	Peace,	शिकायत	*shikāyata*	Complaint
दोस्त	*dosta*	Friend	डिसकाउंट	*ḍisakāuanṭa*	Discount
कीचड़	*khichaḍa*	Mud	दिलचस्प	*dilachaspa*	Interesting
फीकी	*fīkī*	Dull	पीड़ा	*pīḍaā*	Pain
उलझना	*ulazanā*	Tangle	पुराना	*purānā*	Old
कमज़ोर	*kamajaora*	Weak	ढीली	*ḍhīlī*	Loose
दर्द	*darda*	Pain	उबाउ	*ubāu*	Boring
पात्र	*pātra*	Character	भीड़	*bhīḍ*	Crowd
खरा	*kharā*	Pure	खर्चा	*kharchā*	Cost
नयापन	*nayāpana*	Newness	अलग	*alaga*	Different
प्रेम	*prema*	Love	खास	*khāsa*	Interesting
शोर	*shora*	Noise	अनोखी	*anokhī*	Unique
चिल्लाना	*chillānā*	Shouting	खरीद	*kharīda*	Purchase
अनूठा	*anūṭhā*	Unique	प्रतिलिपि	*pratilipi*	copy
ताज़ातरीन	*tājaātarīna*	Fresh	निराश	*nirāsha*	Disappoint
नकल	*nakala*	Copy			

Acknowledgements Authors are thankful to Mr. Mohit Jain, Vardhman Classes for providing language editing support.

Author contributions M.J conceived and designed the study, acquired and analyzed the data, and drafted the manuscript. A. J contributed to the data collection analysis and provided critical revisions to the manuscript. M. J and A. J contributed to the study's design, provided substantial intellectual input, and reviewed and edited the manuscript. M. J contributed to the data interpretation and critically revised the manuscript. A. J and R. J provided overall guidance for the study, supervised the research, and critically reviewed and edited the manuscript. All authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work's accuracy and integrity.

Funding There is no fund, grant or other support received for this research work.

Data availablity The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

AL-Husan, F. B., Amaugo, A. N., & Idiko, B. (2021). Emerging multi-national enterprise (EMNE)'s approaches and challenges for the management of human resources in BRICs Countries. In *International HRM and development in emerging market multinationals*, pp. 314–329.

Arora, P., Bakliwal, A., & Varma, V. (2012). Hindi subjective lexicon generation using wordnet graph traversal. *International Journal of Computational Linguistics and Applications, 3*(1), 25–39.

Bakliwal, A., Arora, P., & Varma, V. (2012). Hindi subjective lexicon: A lexical resource for hindi adjective polarity classification. In *Proceedings of the Eighth international conference on language resources and evaluation (LREC'12)*, pp. 1189–1196.

Balamurali, A. R., Joshi, A., & Bhattacharyya, P. (2012). Cross-lingual sentiment analysis for Indian languages using linked wordnets. *Proceedings of Coling, 2012*, 73–82.

Bose, R., Dey, R. K., Roy, S, & Sarddar, D. (2019). Analyzing political sentiment using Twitter data. In *Information and communication technology for intelligent systems*, (pp. 427–436). Springer, Singapore.

Cambria, E., Hussain, A., Cambria, E., & Hussain, A. (2015). SenticNet. In *Sentic computing: a common-sense-based framework for concept-level sentiment analysis*, pp. 23–71.

Das, A., & Bandyopadhyay, S. (2010). SentiWordNet for Indian languages. In *The 8th workshop on asian language resources (ALR), COLING 2010*, Beijing, China, pp. 56–63.

Esuli, A., & Sebastiani, F. (2007). SentiWordNet: A high-coverage lexical resource for opinion mining. *Evaluation, 17*(1), 26.

Gaikwad, V., & Haribhakta, Y. (2020). Adaptive glove and fasttext model for hindi word embeddings. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pp. 175–179

Garg, K. (2020). Sentiment analysis of Indian PM's "Mann Ki Baat." *International Journal of Information Technology, 12*(1), 37–48.

Garg, K., & Lobiyal, D. K. (2020). Hindi EmotionNet: A scalable emotion lexicon for sentiment classification of Hindi text. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP), 19*(4), 1–35.

Gupta, V., Jain, N., Shubham, S., Madan, A., Chaudhary, A., & Xin, Q. (2021). Toward integrated CNN-based sentiment analysis of tweets for scarce-resource language: Hindi. *Transactions on Asian and Low-Resource Language Information Processing, 20*(5), 1–23.

Hussaini, F., Padmaja, S., & Sameen, S. (2018). Score-based sentiment analysis of book reviews in Hindi language. *International Journal on Natural Language Computing, 7*(5), 115–127.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey. *Multimedia Tools and Applications, 78*, 15169–15211.

Jha, V., Manjunath, N., Shenoy, P. D., & Venugopal, K. R. (2015). HSAS: Hindi subjectivity analysis system. In *Proceedings of the 2015 Annual IEEE India Conference (INDICON)*. (pp. 1–6), IEEE.

Jha, V., Manjunath, N., Deepa Shenoy, P., & Venugopal, K. R. (2016). Sentiment analysis in a resource scarce language: Hindi. *International Journal of Scientific and Engineering Research, 7*(9), 968–980.

Jha, V., Savitha, R., Shenoy, P. D., Venugopal, K. R., & Sangaiah, A. K. (2018). A novel sentiment aware dictionary for multi-domain sentiment classification. *Computers & Electrical Engineering, 69*, 585–597.

Joshi, A., Balamurali, A. R., & Bhattacharyya, P. (2010). A fall-back strategy for sentiment analysis in hindi: a case study. In *Proceedings of the 8th ICON*.

Khanuja, S., Bansal, D., Mehtani, S., Khosla, S., Dey, A., Gopalan, B., & Margam, D. K. et al. Muril: multilingual representations for indian languages. arXiv preprint http://arxiv.org/abs/2103.10730 (2021).

Kumar, S., Kumar, S., Kanojia, D., & Bhattacharyya, P. (2020). A Passage to India": Pre-trained Word Embeddings for Indian Languages. In *Proceedings of the 1st joint workshop on spoken language technologies for under-resourced languages (SLTU) and Collaboration and computing for under-resourced languages (CCURL)*, pp. 352–357

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093–1113.

Mishra, D., Venugopalan, M., & Gupta, D. (2016). Context specific Lexicon for Hindi reviews. *Procedia Computer Science, 93*, 554–563.

Mittal, N., Agarwal, B., Chouhan, G., Bania, N., & Pareek, P. (2013). Sentiment analysis of hindi reviews based on negation and discourse relation. In *Proceedings of the 11th workshop on Asian language resources*, pp. 45–50.

Modi, D., & Nain, N. (2016). Part-of-speech tagging of Hindi corpus using rule-basedmethod. In *Proceedings of the international conference on recent cognizance in wireless communication & image processing*, (241–247), Springer.

Mohammad, S. M., & Turney, P. D. (2013). Nrc emotion lexicon. *National Research Council, Canada, 2*, 234.

Mohanty, A. K. (2008). Multilingual education in India. *Encyclopedia of Language and Education, 5*, 165–174.

Pathak, A., Kumar, S., Roy, P. P., & Kim, B. G. (2021). Aspect-based sentiment analysis in Hindi language by ensembling pre-trained mBERT models. *Electronics, 10*(21), 2641.

Pravalika, A., Oza, V., Meghana, N. P., & Sowmya K. S. Domain-specific sentiment analysis approaches for code-mixed social network data. In *2017 8th international*

Sazzed, S. (2021). BengSentiLex and BengSwearLex: Creating lexicons for sentiment analysis and profanity detection in low-resource Bengali language. *PeerJ Computer Science, 7*, e681.

Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). 2020. Measuring news sentiment. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2020.07.053

Sharma, R., Bhattacharyya, P. (2014). A sentiment analyzer for Hindi using Hindi senti lexicon. In *Proceedings of the 11th international conference on natural language processing*.

Sharma, Y., Mangat, V., & Kaur, M. (2015). A practical approach to sentiment analysis of Hindi tweets. In *Proceedings of the 1st international conference on next generation computing technologies*. pp. 677–680.

Sharma, R., Nigam, S., & Jain, R. (2014). Polarity detection movie reviews in Hindi language. *International Journal on Computational Sciences & Applications (IJCSA), 4*, 49–57.

Sivasankar, E., Krishnakumari, K., & Balasubramanian, P. (2021a). An enhanced sentiment dictionary for domain adaptation with multi-domain dataset in Tamil language (ESD-DA). *Soft Computing, 25*, 3697–3711.

Sivasankar, E., Krishnakumari, K., & Balasubramanian, P. (2021b). An enhanced sentiment dictionary for domain adaptation with multi-domain dataset in Tamil language (ESD-DA). *Soft Computing, 25*(5), 3697–3711.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics, 37*(2), 267–307.

Taher, S. A., Akhter, K. A., & Hasan, K. A. (2018). N-gram based sentiment mining for bangla text using support vector machine. In *2018 international conference on Bangla speech and language processing (ICBSLP)* (pp. 1–5). IEEE.

Thelwall, M. (2017). The Heart and soul of the web? Sentiment strength detection in the social web with SentiStrength. In Cyberemotions, (pp. 119–134). Springer, Cham.

VB, S., Choudhury, M., Bali, K., Dasgupta, T., & Basu, A. (2010). Resource creation for training and testing of transliteration systems for indian languages. In *Proceedings of the Seventh international conference on language resources and evaluation* (*LREC'10*).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/382073040

Ecological Representation in the Maithili Songs of Vidyapati

Article · July 2024

CITATIONS
0

READS
15

2 authors:



Mallika Tosha
Delhi Technological University
10 PUBLICATIONS **5** CITATIONS
SEE PROFILE



Rajiv Ranjan Dwivedi
Delhi Technological University
19 PUBLICATIONS **8** CITATIONS
SEE PROFILE

All content following this page was uploaded by Mallika Tosha on 08 July 2024.

The user has requested enhancement of the downloaded file.

Ecological Representation in the Maithili Songs of Vidyapati

Mallika Tosha
Research Scholar
Department of Humanities
Delhi Technological University

Dr. Rajiv Ranjan Dwivedi
Department of Humanities
Delhi Technological University

Abstract :

Vidyapati (1360-1470 AD) is considered as one of the greatest writers of Mithilanchal, Bihar. He composed songs in such a manner that they became enshrined in the hearts of the natives of Mithila region. He wrote in Sanskrit and Maithili. Still, many of his Maithili songs are sung as folksongs. The present study revealed that his Maithili songs are abundant with the description of the ecological elements. His style contains the aura of the deities through the establishment of their relation with the environment. In his songs, both Gods and human beings are affected by nature. The physical environment acts as a stimulus to their behaviour. The relationship between the compositions of Vidyapati and the environment has been analysed with an ecocritical perspective in this research.

Keywords :

Ecocriticism, Maithili, Environment, Folksongs, Mithila

Introduction :

Vidyapati (1380-1460 AD) was born into an affluent and intellectual family. Also called as *Maithil Kokil* (The cuckoo bird of Mithila), his creative abode is highly related to the folk life of Mithila. The writer created several Maithili compositions, including devotional songs on Hindu deities like Vishnu and Shiva. The simple fact that Shiva is the sole Hindu god whose love and worship have endured over time is why he created so many songs dedicated to him. They have been approved by the straits for any man and woman, regardless of caste or gender. Music dedicated to Shiva were the only devotional songs that could connect with ordinary men and women, regardless of caste or sex, for whom Vidypati penned the songs. *Kirtipataka, Purusparikshа*, *Goraksavijaya* and *Bhuparikrama* are some of his compositions. His compositions were an amalgamation of folk life and Hindu mythology. Maithili folksongs are often not anonymous. Vidyapati composed Maithili songs that are still sung as folksongs. Most commonly, the last line of the lyrics mentions the name of its composer too. Maithili folksongs are abundant with the mention of ecological elements. Vidyapati

(1350-1450) composed songs enriched with the inclusion of flora and fauna of Bihar. According to Pankaj Jha (2018) Vidyapati has been interestingly memorialised and mythologized throughout the course of the previous six centuries or more. His followers believe that he has received supernatural favours and worked miracles in ways that would be outside the scope of a "rational" approach that is fact-focused. The most widely accepted mythology among them states that Lord Œiva was so inspired by the poet's piety that he assumed the identity of a poor man and called himself Ugna in order to serve as the poet's personal attendant. This "episode" in Vidyapati's life is recounted in almost every one of his numerous biographies. Any discussion about him with a Maithil, the author claims, must mention the story.

Discussion :

The anthology *Vidyapati Rachnavali* by Dubey (1981) contains various verses of the Maithili poet Vidyapati. Dubey has categorized these in six categories. The categories are *upeksha geet* (songs of hope), *milan-gopan geet* (songs of love), *har gouri geet* (songs of Shiva and Gouri), *vandana geet* (Songs of devotion), *ritu geet* (Seasonal songs) and *samanya geet* (simple songs). These songs reflect on the cultural, environmental, philosophical and sociological aspects of the people who are singing and their surroundings. In 1971, Ramanath Jha wrote about Maithili:

> The State of Maithili publications in 1970 was as depressing as during the previous years. There has been no change in the apathetic attitude of the Government of Bihar towards Maithili but much more depressing is the apathy of the people speaking the language. Even the educated among the speakers of this language do not purchase Maithili books and Maithili publication, therefore, is commercially a losing concern (Ramanath Jha, 1971, p. 86).

Those were the times when the folk languages started depleting. Though the Indian government takes several initiatives in contemporary times to save the folk languages, but their extinction seems inevitable. The eminence of this language before could be seen in the writings of Vidyapati. According to Mishra (1966), his songs are embodied with a touch of Mithila folk life. He further comments :

Armed with a thorough knowledge of Sanskrit poetics, he does not hesitate to bring together the sacred Shastras and the profane elements of the Maithili folklore, like the myth of the twin rivers Ganges and Jamuna. Some Folk-songs of Mithila can be easily traced in the Pada vali: The lonely, melancholy girl, separated from her lover, pleads with the Kokila bird :

'The moon spits fire,

lotuses droop and loaded with fragrance mingle in sad love.

Kokila, bird of spring, why do you torture me?

Why do you sing your love-provoking song?' (Mishra, 1966, p.136)

This song of Vidyapati, quoted by Mishra, is full of nature imagery. The inclusion of words such as "lotuses," "*kokila,*" and "spring" all coalesce together to offer the readers synaesthesia.

In one of the compositions, Vidyapati depicts the pang of a wife whose husband has gone to another region:

Mor Piya sakhi gel door des
Jouvan daye gel saal sanes
Maas asadh unat nav megh

(Vidyapati, 2018, p. 271)

O friend, my husband has gone to a distant land

My youth has given me a message this year

The season is rainy, and new year's heavy rain is there.

The ongoing of season has been equated with the conditions of the woman who is full of sorrow as her husband is not with her. Heavy rain is symbolic of the tears that the woman is shedding. This description points to the innate nature of human beings to see their psyche in the various happenings of their natural environment.

Vidyapati applies vivid imagery in a consistent manner. There are several instances in the creative expression of Vidyapati that hint at the essence of nature and glorify the ecosystem along with its flora and fauna. One of the folksong runs as follows:

Lata taruvar mandap dia, nirmal sasdhar bhiti dhablih

Parun naaal aipan bhal bhel, raat Pallav nav pahiran del

Gabah mai he mangal aaye, basant-bibah kanan thali jaye

Madhukar-ramani mangal gab, dujbar kokil mantra padhab

Karu makrand hathodal neer, bidhu bariyati dheer Sameer

Kanau keaasuti toran tool, lava bithrak belik fool

Kesu kusum karu siddur daan, joutuk paol maanini maan

*(*Vidyapati, 2018, p.176)

The tree embellished with climbers got venerated for the wedding, the pious moon cleaned its walls.

The lotus lot added to the beauty, the new tender leaves clothed the night

O maids, sing an auspicious song together, and witness the wedding of Spring

Bumblebee sings joyful song, the brahmin birds are making the cuckoo recite mantras.

The male voice acted as promising pious water, slow wind turned groom's side group

Plumeria and banana turned decorative strings; the scattered jasmine flowers became auspicious grains

Flame of the forest put vermillion, and the groom acquired respect as brides' gift.

There is a recurrent inclusion of the seasonal changes of nature in the compositions of Vidyapati. In these texts, the reflection of the seasons includes a plethora of personification and symbolism. One of his texts read as following:

Magh maas siri Panchami ganjaili, navam maaa pancham haruaai hey

Ati ghan peera dukh bar paaol, banaspati bheli dhai hey

Subh khan bera sukul pakkh hey, Dinkar udit samai he

Sorah sampun batis lakhne sah, janam lel riturai hey

Nachaye jubitjana harkhit man, janmal baal madhai hey

Madhur maharas mangal gaabaye, manini maan urai hey

Bah malyanil ot uchit hey, nav ghan bhau ujiaara

(Vidyapati, 2018, p.172)

(The weather is again about to beget a son, those nine months were painful

She was full of sadness, so all the herbs came together like her midwife

In pious timing, spring was born, complete with abundant virtues

On this occasion of baby spring's birth, all young ladies danced

Egoistic women shed their egos and started singing

The Malaya Mountain breeze touched the baby by bringing rain

The spring flowers were mesmerising, so they were used in making a cradle)

This folk utterance of Vidyapati is connotative of the effects that nature has on its inhabitants. When the poet says that all women had an impact of Spring on them, it exhibits the effect that nature could have on our minds. In the past, several established authors such as William Wordsworth, John Keats, Percy Bysshe Shelley, and Henry David Thoreau have asserted the force that nature has on driving our emotions. Pankaj Jha puts forth :

> A literary device that helped Vidyapati accomplish this was the tradition of 'subhāsita' in Sanskrit literature. In fact, the whole text of Purucaparīksā, and occasionally even Likhanāvalī, is liberally sprinkled with independent couplets of (un) identifiable origin. These were drawn from the free-floating reserve of an ever-growing śloka tradition, called subhāsita, literally, 'well said' a 'witty saying' or an aphorism. As such subhāsita belonged both to the folk as well as to the classical tradition, and very often, to a classical rendering of folk traditions (Pankaj Jha, 2018, p. 176).

Through the utilization of such proper tropes, Vidyapati enhanced folksongs and brought them from the rustic sphere to the zone of intellectuals.

Conclusion :

There lies a great inclusion of various environmental elements in the Maithili songs of Vidyapati. The reassertion of Indian folk values could lead to a better future in terms of the environment. By reclaiming the past lineage of folk beliefs and its positive outlook towards the ecosystem, we can guide ourselves on how to conserve biodiversity and do less harm to the ecosystem. The folk songs of Vidyapati acts as a successful medium in judging the psyche of human beings when they connect themselves with the environment and knowingly or unknowingly disrupt nature. Equating the connection between human beings and the environment with the reflections in the compositions of Vidyapati, this research intends to offer a newer perspective on understanding the human-nature relationship. Through such examination of mindset, researchers could develop sensitivity and raise the oneness of the native with the ecosystem in which they live.

References :

1. Dubey, A. N. (1981). *Vidyapati Rachnavali.* Shakti Publication.
2. Jha, P. (2018). *A Political History of Literature: Vidyapati and the Fifteenth Century*. Oxford University Press. p. 176.
3. Jha, R. (1971). Maithili: Results in Spite of Handicaps. *Indian Literature*, *14* (4), 86-91.
4. Misra, L. P. (1966). A Critical Note on Vidyapati. *East and West*, *16* (1/2), 126-138.
5. Vidyapati (2018). *Vidyapati Padavali* (Chaturvedi, R. P. & Tiwari A. Ed.). Sahitya Sarovar Publication. pp. 172,176, 271. (Original Work Published 1400- 1480).
6. Note: All translations, unless mentioned are the works of the corresponding author.

View publication stats

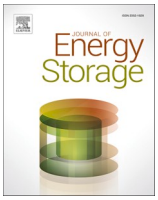
Journal of Energy Storage 97 (2024) 112938



Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est



Review Article



Effect of various energy storage phase change materials (PCMs) and nano-enhanced PCMs on the performance of solar stills: A review

Aseem Dubey, Akhilesh Arora *

Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Keywords:
Solar still
PCMs
NePCMs
Hybrid PCMs
Bioactive PCMs
Productivity

ABSTRACT

The shortage of potable water is a vital issue for isolated communities in some countries. Utilizing solar still (SS) has proven to be an effective and economical method for producing potable water, but it suffers from low productivity. Among the multiple innovative techniques employed to enhance the performance, employing phase change materials (PCMs) ensures enhanced incessant production of potable water during the night, thus bridging the gap between supply and demand. Further, as an emerging field, the application of nano-enhanced PCMs (NePCMs) has been found promising in augmenting the thermal conductivity of PCMs. In this manuscript, a review of the recent studies on the performance enhancement of solar stills incorporated with PCMs, nano-enhanced PCMs, hybrid PCMs, bioactive PCMs, etc., has been carried out. The novel approaches have demonstrated the enhancement of yield in the range of 31–100 % and 49.5–132 % by incorporating PCMs and NePCMs, respectively, using passive solar stills, while reaching a maximum of ~140 % and ~ 244 % with active solar stills. Bioactive NePCM has shown enhanced yield by ~124.2 % compared to conventional solar still. The use of hybrid PCMs and NePCMs is found attractive and enhances the yield further by ~9 % compared to single PCM. It is further revealed that adding sensible heat storage materials with basin water as a hybrid approach with PCMs/NePCMs loaded solar stills is promising, along with augmenting the performance using magnetic field and UV plasmon resonance. This review also emphasizes the opportunity and future research and would form a road map for the researchers to explore appropriate hybrid combinations to compete with the other technologies.

1. Introduction

Potable water is one of the basic human needs and faces global challenges due to its increasing scarcity and non-sustainable supply in different regions globally. Around one billion people suffer from freshwater scarcity, which will continue to increase due to population and industrial growth [1]. A freshwater scarcity of~40–80 % has been reported in China, India, and the United States [2]. The water scarcity score and the daily available solar energy in some countries are depicted in Fig. 1. Around 17 nations in the Middle East, North Africa, and Sub-Saharan Africa are severely water-stressed. The report published in the year 2018 has mentioned the top five countries facing water scarcity based on their population, i.e., India (163.1 million), Ethiopia (60.5 million), Nigeria (59.4 million), China (57.4 million), and the Democratic Republic of Congo (46.87 million) [3]. Moreover, most water-scarce countries have ample solar energy potential, which can be effectively used for desalting saline/impure water and assist countries in reaching carbon neutrality targets, reducing carbon emission levels to 50–60 % by 2050 [4,5]. Diverse solar desalination methods include multiple-effect humidification (MEH), multi-stage flash distillation (MSF), multiple-effect distillation (MED), multiple-effect boiling (MEB), humidification–dehumidification (HDH), and solar stills. Numerous studies have been carried out to ensure access to potable water, but these are challenging because of the increased expenses associated with purification, transportation, and storage [6]. The deployment of solar still in the duopoly market would address the scarcity of drinking water in an eco-friendly manner and at an affordable cost, especially for underserved populations in rural and island regions, and hence, would produce more welfare to the community than monopoly technologies and can be an effective solution [7,8]. Nevertheless, low productivity from conventional solar stills is the biggest flaw, which hinders its commercialization, and it is crucial to incorporate modifications to improve its productivity and efficiency [9,10].

Several design modifications have been accomplished to improve the productivity, efficiency, and economic viability of solar stills, which have been historically documented since the 1970s, classifying them as passive and active solar stills. To enhance the productivity of solar stills,

* Corresponding author.
E-mail address: akhilesharora@dce.ac.in (A. Arora).

https://doi.org/10.1016/j.est.2024.112938
Received 14 January 2024; Received in revised form 19 June 2024; Accepted 8 July 2024
Available online 25 July 2024
2352-152X/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Nomenclature

A_b	area of basin (m^2)
A_g	area of condensing cover (m^2)
B	Boltzman constant (1.381×10^{-23} J/K)
C_w	specific heat capacity (J/kgK)
d_{np}	diameter of nanoparticle (m)
h_{bp}	overall h.t.c. from basin to PCM (W/m^2K)
h_{bw}	convective h.t.c. from absorber to basin fluid (W/m^2K)
h_{1w}	total internal h.t.c. (W/m^2K)
h_{kg}	h.t.c. of inner glass covers (W/m^2K)
ho	top h.t.c. from glass surface (W/m^2K)
h_{go}	overall top h.t.c. from outer surface (W/m^2K)
$I_s(t)$	radiations (W/m^2)
K_b	basin's thermal conductivity (W/mK)
K_{pcm}	basin's thermal conductivity (W/mK)
l_b	thickness of basin (m)
lg	thickness of glass (m)
l_{inc}	thickness of insulation (m)
l_{pcm}	PCM thickness (m)
M_w	water mass (kg)
M_{pcm}	mass of PCM (kg)
Pr	Prandtl number
Gr	Grashof number
Q_{char}	heat flux during charging (W/m^2)
Q_{dis}	heat flux during discharging (W/m^2)
t	time (s)
T	temperature (°C)
∂t	incremental temperature rise (°C)
U_{ins}	overall heat loss coefficient from PCM (W/m^2K)
V_a	speed of wind (m/s)
$h.t.c.$	heat transfer coefficient

Greek symbols

$\propto'$	absorptivity
$\varnothing$	fractional nanomaterial
μ	viscosity (NS/m^2)
ϵ	correction factor in the brownian motion term

Subscripts

g	glass
b	basin
w	water
i	initial
ins	insulation
equ	equivalent
s	solid
l	liquid
m	melting
np	nanoparticle

various initiatives, such as enhancing the solar collection, increasing basin water temperature, reducing heat losses, and re-utilization wastage heat, can be implemented [11]. The various modified designs and new enhanced techniques, such as the use of insulation [12], optimum cover tilt [13], external condensers [14,15], solar collectors/concentrated parabolic reflectors [16–18], V-groove basin [19–21], waste heat recovery of PV/T system [22,23], photothermal absorbers [24], glass cover cooling [25–28], thermoelectric cooling [25,29], wind ventilators [30], reflectors [10], domed cover designs [31,32], prismatic absorber basin [6], pulsating heat pipe [33,34], heat storage materials [35,36], biomaterials [37], porous materials [38–40], Sensible heat storage materials [41] nanomaterials [42–45] and nano-enhanced PCM [32,46], and their hybrid applications [47] have been researched. At optimal operating conditions, an optimal yield of 3.10 kg/m^2 from a dual slope SS, while 5.0 kg/m^2 from a single slope SS were reported [48,49]. The mean energy efficiency of ~30 % has been reported using CSS, reaching up to ~60 % under idealistic conditions [50]. Moreover, integrating the evacuated tube heat pipe and external condenser with solar still has enhanced the daily yield by 2.13 times [51]. A detailed overview has been presented for the solar stills integrated with heat pipes and pulsating heat pipes, concluding the lowest production cost of distillate when coupled with an evacuated heat pipe [52].

Thermal energy storage systems (TES) have been the emerging waste/excess heat recovery methods, potentially boosting energy resilience by storing thermal energy. The TES are classified as sensible heat storage, latent heat storage, and thermochemical energy storage systems, which have been extensively reviewed [53]. Spherical rock salt balls (0.50, 1.0, 1.50, and 2.0 cm diameter) as a low-cost sensible energy storage material has been investigated on hemispherical solar still and found promising [41]. They have reported enhanced daily productivity

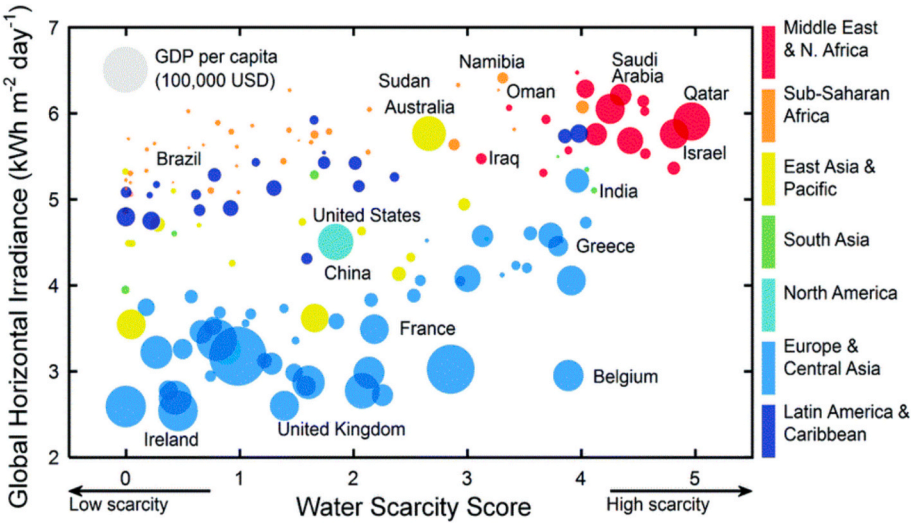


Fig. 1. Water scarcity score and availability of daily solar radiation [4].

of hemispheric solar still, reaching 6.77, 6.25, 5.92, and 5.67 kg/ m^2 by incorporating ball sizes of 2.0, 1.50, 1.0, and 0.50 cm, respectively, compared to 4.65 kg/m^2 yielded by the reference solar still. Moreover, using the spherical salt balls of respective sizes decreased the yield by 8.83 %, 23.47 %, 19.73 %, and 16.84 % compared to reference solar still. From the summarized study, it has been explored that much research has been carried out on improving the performance of the solar still using different wicked materials such as jute, linen, cotton wicks and PCMs and NePCMs as well as combinations to ameliorate the productivity of solar still. Incorporating PCMs can be beneficial for storing the excess heat, enabling the system to operate beyond daylight and thus increase the system reliability and enhanced yield by 49 % was reported using paraffin compared to conventional solar still [54]. Radhwan [55] has fabricated five stepped basins solar still for heating and humidification of agricultural greenhouse. Each basin was placed on a PCM-filled slab (paraffin 0.01–0.03 m thick), insulated from the bottom using a foam of 0.02 m thickness. The re-circulated air from the greenhouse enters the still from the bottom, flows between the basins and glass cover, and after heating and humidification it flows back into the greenhouse. The total yield of 4.6 kg/m^2 with 57 % energy efficiency with PCM, while 4.9 kg/m^2 with 61 % efficiency without PCM has been obtained, maintaining 0.015 m water depth and 0.001 kg/s airflow rate. Furthermore, a single slope solar still of GI basin (1 m^2), insulated with 0.05 m thick foam, and v-corrugated basin with Cu absorber (0.002 m thick) and incorporated with 18 kg paraffin wax beneath was tested with and without the PCM. The glass cover was 0.003 m thick and inclined at 30°. Thirteen copper tubes of diameter 0.01 m were inserted through the PCM to counter the problem of the increasing volume of wax upon melting. Adding wicks over the corrugated absorber with PCM was also investigated and found an increase in daily yield by 12 % and 11.7 % using PCM and PCM/wicks arrangement, respectively, compared to the still without PCM for 25 kg basin water [56]. Using PCM, day yield was decreased by 7.4 %, while night yield increased by 72.7 %. Moreover, the cost of distillate production has been estimated as 0.08369, 0.09558, and 0.07182 $/kg using PCM, PCM with wicks, and without PCM solar still, respectively. Kabeel and Abdelgaied [57] studied the effect of PW on single slope solar still made of 0.0015 m thick GI sheet of 0.072 m^2 basin area, insulated by 0.005 m thick fiber glass. The glass cover was 0.003 m thick and inclined at 30°. The daily yields of 7.54 kg/m^2 and 4.51 kg/m^2 were obtained from the solar still with PCM and without PCM. The production cost of distillate using PCM was found to be lower using PCM (0.03 $/kg) compared to still without PCM (0.032$/kg), due higher productivity. In addition to experimental analysis, numerical studies including thermal modeling and computational fluid dynamics (CFD) simulations have also been carried out to study the behavior of energy storage materials. El-Sebaii et al. [58] have developed a mathematical model to study the impact of PCMs using stearic acid in conventional solar still and found a higher yield by 80 %. The effect of PW on a dual pass solar air heater–coupled conventional solar still with hot air injection has shown an enhanced yield of 9.36 kg/m^2, which is ~108 % higher than the yield obtained from conventional solar still (4.5 kg/m^2) [59]. Khalilmoghadam et al. [60] have reported daily yield and production cost of distillate as 6.3 kg/m^2 and 0.0093 $/kg, respectively, using acetone as PCM on pulsating heat pipe solar still. Pulsating heat pipes (PHPs) were used in the thermal storage system as heat transfer devices between PCM and basin water due to their high conductivity. To prevent heat transfer from the basin to the PCM during daytime, the basin was placed in the center of the chamber and away from the PCM to avoid direct contact with it. A numerical study has been conducted on organic and inorganic PCMs to investigate productivity, retention time, and economics, and found that CSS incorporated with A-48 PCM yielded maximum with lower production cost and was recommended for use from the environmental aspects [61]. Hafs et al. [62] have carried out a numerical simulation of passive and parabolic trough collector integrated active solar still with corrugated absorber surface (rectangular, triangular and spherical ripples) embedded with PCM. The results have shown that the passive solar still with rectangular ripples yields higher than other configurations with increased productivity by 109 % and 42 % than the conventional solar still and PCM-loaded CSS, respectively. Moreover, PTC integrated solar still with rectangular ripples has improved the yield by 549.36 % compared to passive mode of operation. Research initiatives incorporating PCMs have inferred higher productivity from the solar stills. Moreover, PCMs are associated with the discrepancy of low thermal conductivity [63].

To overcome the discrepancy of low thermal conductivity, the surface area of PCMs/water can be enhanced by using extended surfaces, dispersing nanoparticles, or both [64,65]. Karthikeyan et al. [66] tested solar stills containing pebbles, pebbles with color dye, gravels, gravels with color dye, a glass ball, a glass ball with color dye, and Al_2O_3 nanoparticles and found maximum productivity of 6.3 kg/m^2 using Al_2O_3 nanoparticles. The thermal conductivity of paraffin wax (PW) has been improved by four times by mixing alumina nanoparticles (5–10 wt %), which was enhanced further to 35–40 % by adding carbon nanotubes (1–2 %) [67,68]. Moreover, mixing alumina and graphite nanoparticles into myristic acid (PCM) enhanced the thermal conductivity by 125 % [69]. Incorporating copper nano-composite with PW enhanced the yield by ~125 % [70], decreased melting point by 15 % using CuO+ PW blend [71], and enhanced energy storage capacity with graphite-based PW [72] have been reported. The thermal conductivity of the PCM has been increased by over 200 % by the hybridization of Al_2O_3-CuO in the ratio of 3:1 [1]. A comparative experimental study of the stepped still using two types of basin liners (curved and corrugated), cloth wicks, CuO/PW, and vapor suction has revealed that the use of corrugated liner with wick, with wick + CuO/PW, and with wick + CuO/PW + vapor suction improved the yield by 95 %, 127 %, and 170 %, respectively, compared to conventional solar still [73]. Implementing the stepped absorber of multi-conical trays with fountain feed seawater supply enhanced the absorption surface area. A dome-shaped solar distiller with a stepped conical absorbent basin with multi-trays was tested by incorporating a blend of PCM/CuO nanoparticles (NePCM) with a fountain feed supply and found significant enhancement in yield (7.21 kg/m^2) [32]. Saeed et al. [74] designed a corrugated drum solar still, combining PCM with silver nanomaterial and nano-painting in the basin. The effect of rotating the drum at various speeds (0.05, 0.2, 0.3, 0.5, 0.7 rpm) was also studied. The result revealed an increased yield of 318 % at 0.1 rpm compared to conventional solar still. The numerical study on NePCMs loaded solar still inferred that the TiO_2 + PW blend is the most suitable and has stable thermophysical properties during sunshine hours, improving the daily yield to 6.6 kg/m^2 [75]. Numerical studies using PW and Al_2O_3 + PW have been carried out on COMSOL 5.3 software to optimize the performance of CSS, and results show that 3 vol % of Al_2O_3 nanoparticles with 1 kg of PW improve the daily yield by 20 %. [76]. Moreno et al. [77] explored RT70 HC as an improved PCM, predicting a day/night yield of 4.14/1.28 kg/m^2 for 0.01 m water depth. Shoeibi et al. [78] numerically investigated the use of hybrid Al2O3/TiO2 water nanofluid to cool the condensing cover of solar still, predicting increased yield by 11.09 %. Jamil et al. [79] have reviewed the application of advanced energy-stored materials in direct solar desalination and reported a significant enhancement in the performance of their applications. Recently, Hemmatian et al. [80] positioned the wax within the thermosyphon/pulsating heat pipe evacuated tube at 50 % and 100 % filling ratios and found enhanced daily yield and energy efficiency by 40.7 and 42.9 %, respectively, but simultaneously the production cost of distillate was increased by 19.5 % compared to CSS. Elsheikh et al. [6] have identified artificial neural network methods (ANNMs) as a tool for simulation for better decisions and found underexplored forecasting opportunities for the operational production of solar distillation technologies, which can provide timely feedback to decision-makers.

Conventional passive solar still has not yet been commercialized because of its lower productivity. To solve the issue of low productivity, applications of PCMs and NePCMs are emerging, including hybrid

nanomaterials and bioactive-driven materials for the justified deployment for commercialization [7]. Fig. 2 depicts the trend of publications in the field of solar still focusing on the application of PCM and NePCM, and shows an increment during the last five years. This trend signifies a continuous research interest globally in the field (data searched as keywords on Web of Science).

This work aims to review the incorporation of novel materials in recent studies that can improve potable water productivity by solar stills. Therefore, this article focuses on the impact of varied energy-stored materials and methods that can be deemed effective for enhancing the performance of solar stills due to their modified melting points and thermal conductivities, such as applications of hybrid PCMs, diversified NePCMs, and their hybrid combinations. The present work has a novelty that is reflected by the comprehensive coverage of enhanced configurations of solar stills such as dome-shaped with fountain feed, hybrid PV/T, and simultaneously incorporation of different energy storage materials such as bioactive nanomaterials, rehashed edible oils, hybrid NePCM, and influence of magnetic field being done in recent time, which were not included before according to the knowledge of the authors. This updated review provides a unique study for the future researchers, exploring new materials/combinations to investigate diverse novel configurations of solar stills to improve the performance. This review will facilitate a better understanding of the progress and serve as a road map for the commercialization of solar still. Additionally, the future possible works in this area are presented.

The article is categorized and structured under different sections. Section 2 describes the energy storage materials, Section 3 explains the thermal analysis for each element of the solar still and PCM during phase transformation mode, Section 4 details the studies on PCM incorporated solar stills along with the combinations of sensible heat storage, Section 5 comprises of research configuration concerning to solar stills incorporated with NePCMs, hybrid NePCMs, bioactive nanomaterials, hybrid PCMs, and their combinations, and finally, in Section 6, conclusions and future direction of research and applications are presented.

2. Energy storage materials

The energy storage materials have been categorized as sensible and latent heat storage materials. Sensible heat storage materials have two benefits, namely availability and compatibility. Latent heat storage materials are preferred over sensible heat storage materials due to their larger storage capacity/m3 and energy retaining time [81]. PCMs are classified according to their nature, as shown in Fig. 3. The organics are

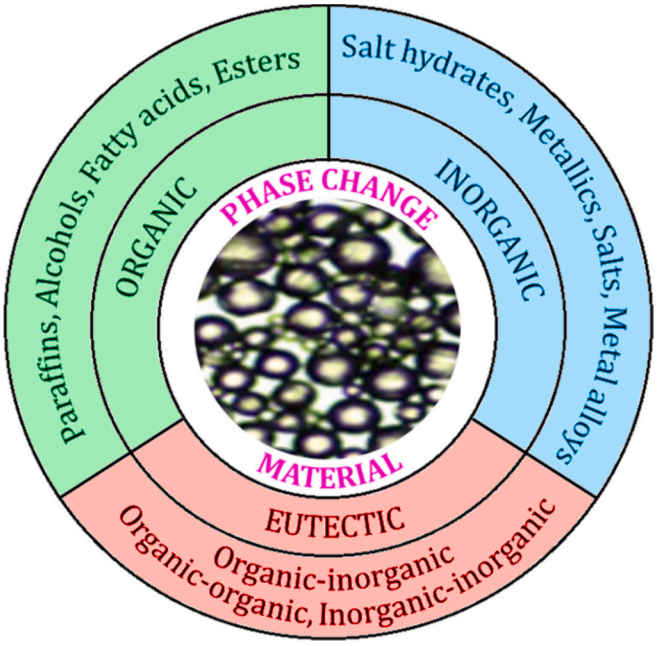


Fig. 3. Classification of PCM.

paraffin or non-paraffins, while inorganic PCMs cover mainly salt hydrates. The melting point of most organic compounds is <80 °C. Eutectics are a mixture of PCMs, and many blends have been investigated to suit the desired applications. Organic PCMs are inexpensive, stable, recyclable, non-corrosive, and possess high latent heat of fusion. However, their main drawbacks are low thermal conductivity, flammability, volatility, and comparatively lower phase change energy [82]. Inorganic salt hydrates are highly conductive than organic PCMs and cheaper but lead to phase segregation, supercooling difficulties, low latent heat, and environmental hazards. The eutectics are the mixture of two or more PCMs, each solidifying and melting concurrently to serve a specific purpose. Eutectic PCMs possess higher heat conductivity and density but are associated with lower specific and latent heat capacity when compared to paraffin/salt hydrates; however, their melting point can be adjusted depending on the mixing ratio [83]. The use of thermochemical energy storage is under the research and development stage.

The appropriate selection of PCM is influenced by phase change

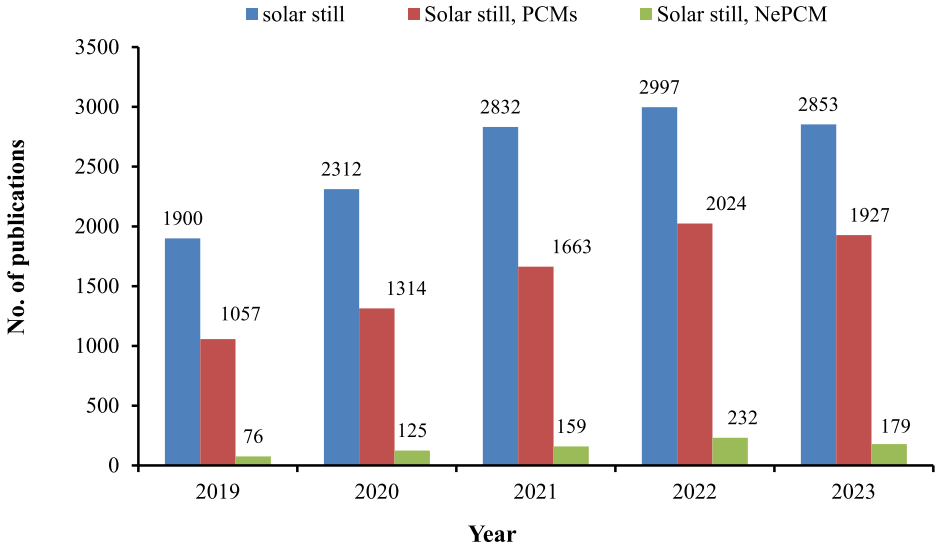


Fig. 2. Publication details with keywords as solar still, solar still with PCM, NePCM.

temperature, latent and specific heat, and their thermal conductivity, and the most suited melting temperature range of PCMs is 40 to 80 °C for solar heating applications [84]. For the passive stills, Su et al. [85] have reported the selection of PCMs of a low melting point range of ~30-45 °C, with preferred materials being paraffin and acetamide. Mousa and Gujarathi [54] have found that PCMs with a melting point range of 40-50 °C are favorable in solar still and yield higher than the PCMs with a melting point below 40 °C. The effect of inorganic PCMs (potassium dichromate, sodium sulfate, and sodium acetate) on the yield of solar still was investigated [86], and sodium sulfate was found promising and economical, however, due to corrosiveness, it requires a stainless steel container [87]. The course of PCM during charging and discharging is shown in Fig. 4. As the radiations increase, PCM temperature increases from point 'a' and reaches the melting point 'b'. At this point, PCM starts melting at constant temperature till point 'c'. After complete melting, upon further heating, the temperature rises to a peak 'd' and further drops back around sunset to point 'e' till it completely solidifies (point f) with the discharge of storage energy and finally reaches near the initial condition during the night (point g).

Nowadays, bio-based PCMs are derived from eco-friendly renewable resources, such as palm kernel oil, palm oil, coconut oil, soybean oil, etc. The wide range of melting points, high latent heat, stability, availability, non-toxicity, and less flammability make bio-based PCMs more attractive than paraffin [84]. The study on various fatty acids has revealed that stearic acid posses maximum heat storage capacity and minimum with capric acid among palmitic, myristic, stearic, capric, and lauric acids. To overcome the possible leakage and diffusion of less viscous PCMs on melting, macro-encapsulation and microencapsulation have been carried out, which also improve the heat storage capacity [88]. Microencapsulation increases the attainment of PCMs by enhancing the thermal transfer area and avoiding leakage issues after melting. In comparison, microencapsulation contains PCM-filled capsules made of silica/organic polymers of size ranging from 1 to 300 μm and reduces toxicity upon mixing liquid and solid PCMs [89]. Zhang et al. [90] have investigated PCM microcapsules with Ti_4O_7 nanomaterial for solar thermal energy storage and found absorptivity up to 88.3 % and photothermal storage efficiency up to 85.36 % using PCMs and SiO_2/Ti_4O_7 compared to pure PCM (24.14 %). The encapsulation makes the PCMs compatible with containers and overcomes their main disadvantages, especially salt hydrates [91]. The selection of PCM for specific applications is based on its latent heat, thermal conductivity, sensible heat, melting point, and chemical and kinetic characteristics.

3. Thermal analysis of solar still loaded with PCM

Fig. 5 shows a diagrammatical presentation of single slope solar still incorporated with PCM. The radiations are passed through the condensing cover and water and further get absorbed by the basin liner.

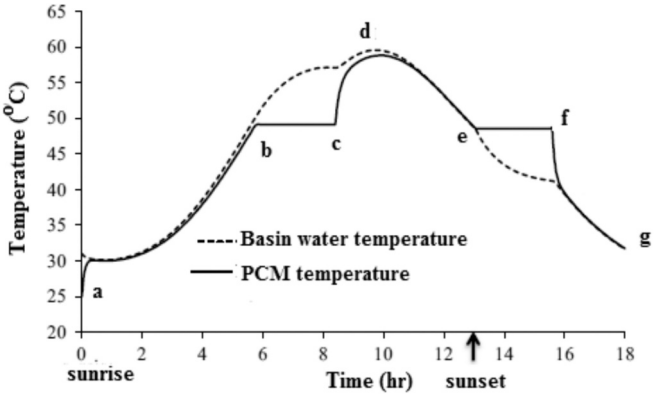


Fig. 4. PCM behavior during charging/discharging.

After absorption, the absorber plate gets heated and transferred heat partially into basin water, while remaining to PCM layer through the bottom. The charging of PCM occurs during radiation availability and decreases the bottom losses. On heating, water evaporates, condenses under the top cover, and trickles down in the channel fabricated with the still body at the lower end. During low/off sunshine hours, the basin liner temperature starts decreasing. As a result, liquid PCM starts discharging heat by transferring it back to the basin absorber and basin water until complete solidification is reached. Thus, the still yields continuously at night, even at lower water depths, because of the increased temperature difference between water and condensing cover.

The energy balance for each component has been formulated using the assumptions presented [92,93].

3.1. Solar still loaded with PCM (charging period)

The energy balance at various components of the system can be expressed as [58,94];

3.1.1. Lower surface of condensing cover

$$\alpha_g' I_s(t) A_g + h_{1w} A_b (T_w - T_{gi}) = h_{kg} A_g (T_{gi} - T_{go}) \tag{1}$$

3.1.2. Outer surface of condensing cover

$$h_{kg} A_g (T_{gi} - T_{go}) = h_o A_g (T_{go} - T_a) \tag{2}$$

3.1.3. Basin liner

$$\alpha_b' I_s(t) A_b = h_{bw} A_b (T_b - T_w) + h_{bp} A_b (T_b - T_{pcm}) + m_b c_b \frac{dT_b}{dt} \tag{3}$$

3.1.4. Energy balance for water mass

$$\alpha_w' I_s(t) A_b + h_{bw} A_b (T_b - T_w) = h_{1w} A_b (T_w - T_{gi}) + M_w C_w \frac{dT_w}{dt} \tag{4}$$

3.1.5. Energy balance at PCM

The energy balance during charging at PCM can be represented as;

$$h_{bp} A_b (T_b - T_{pcm}) = U_{ins} A_b (T_{pcm} - T_a) + M_{equ} \frac{dT_{pcm}}{dt} \tag{5}$$

where M_{equ} is the PCM's equivalent heat capacity and may be evaluated as follow [58];

$$M_{equ} = M_{pcm} C_{s,pcm}, \text{ if } T_b < T_m \tag{6a}$$

$$M_{equ} = M_{pcm} L_{pcm}, \text{ if } T_b = T_m + \partial T \tag{6b}$$

$$M_{equ} = M_{pcm} C_{l,pcm}, \text{ if } T_b > T_m \tag{6c}$$

Solving Eqs. (1–5), and assuming negligible mass/stored energy in basin liner, the following expressions can be obtained;

$$T_{gi} = \frac{\alpha_g' I_s(t) A_g + h_{1w} A_b T_w + h_{go} A_g T_a}{h_{1w} A_b + h_{go} A_g} \tag{7a}$$

$$T_{go} = \frac{h_{kg} T_{gi} + h_o T_a}{h_{kg} + h_o} \tag{7b}$$

$$T_b = \frac{\alpha_b' I_s(t) + h_{bw} T_w + h_{bp} T_{pcm}}{h_{bw} + h_{bp}} \tag{7c}$$

where, $h_{kg} = \frac{K_g}{l_g}, h_{bp} = \frac{K_b}{l_b}$, $U_{ins} = \left[\frac{l_{ins}}{K_{ins}} + \frac{1}{2.8+3V_a}\right]^{-1}, h_o = 5.7 + 3.8V_a$

$h_{go} = \left[\frac{1}{h_{kg}} + \frac{1}{h_o}\right]^{-1}, h_{bw} = \frac{K_w}{l_b}\left[C(GrPr)^{1/n}\right]$

Following [93], the value of C and $1/n$ is reported as 0.54 and 0.25

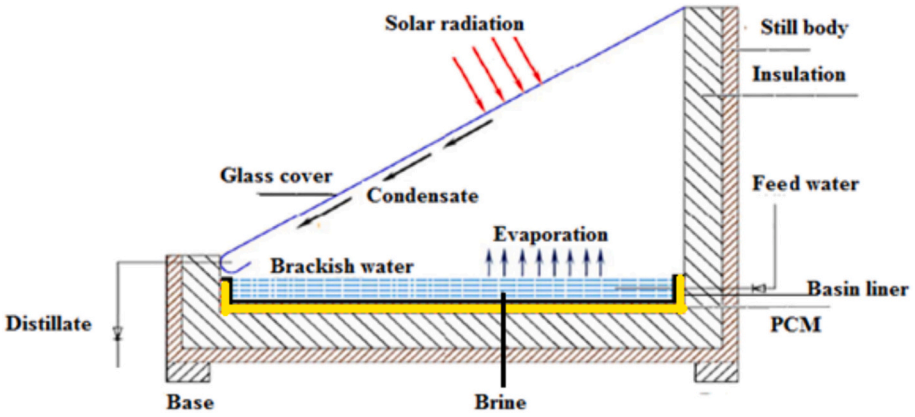


Fig. 5. Diagram of PCM incorporated solar still.

for passive solar still. The heat transfer coefficients can be estimated using the correlations proposed [22]. After rearranging various terms, the water and PCM temperatures can be estimated after time interval 0-t as follows [58,93];

$$T_w = \frac{\overline{f_1(t)}}{a}(1-e^{-at}) + T_{wi}e^{-at} \tag{8}$$

$$T_{pcm} = \frac{\overline{f_2(t)}}{b}(1-e^{-bt}) + T_{pcm,i}e^{-bt} \tag{9}$$

where

$$a = A_b \frac{\left[\frac{h_{bp}h_{bw}}{(h_{bp}+h_{bw})}\right] + \left[\frac{h_{1w}h_o}{(h_{1w}+h_o)}\right]}{M_w C_w}$$

$$b = A_b \frac{\left[\frac{h_{bp}h_{bw}}{(h_{bp}+h_{bw})}\right] + U_{ins}}{M_{equ}}$$

$$f_1(t) = \left(\alpha_w^{'} + \alpha_b^{'}\frac{h_{bw}}{(h_{bp}+h_{bw})}\right) I_s(t)A_b + \frac{\left[\alpha_g^{'} I_s(t)h_{1w}A_g + h_{1w}h_o T_a A_g + h_{bp}h_{bw}T_{pcm}A_b\right]}{(h_{1w}+h_o)}$$

$$f_2(t) = \left(\alpha_{b_}^{'} I_s(t) + h_{bw}T_w\right)\frac{h_{bp}A_b}{(h_{bp}+h_{bw})} + U_{ins}T_a A_b$$

3.2. Discharging mode of PCM

The energy balance equations during low/off sunshine period of solar still when PCM discharges are detailed as under;

3.2.1. Energy balance for PCM during time interval of 'dt' can be written as;

$$M_{equ}\frac{dT_{pcm}}{dt} = \left[h_b^{'}(T_{pcm}-T_b) + U_{ins}(T_{pcm}-T_a)\right]A_b \tag{10}$$

where, $h_b^{'} = \frac{K_{pcm}}{l_{pcm}}$.

3.2.2. Energy balance at basin liner

$$h_b^{'}(T_{pcm}-T_b) = h_{bw}(T_b - T_w) \tag{11}$$

3.2.3. Energy balance at basin water

$$h_{bw}(T_b - T_w)A_b = h_{1w}(T_w - T_{gi})A_b + M_w C_w \frac{dT_w}{dt} \tag{12}$$

Following [58,94], the water and PCM temperatures during discharging mode can be estimated by solving eqns. (11–15).

$$T_w = \frac{\overline{f_3(t)}}{c}(1-e^{-ct}) + T_{wdi}e^{-ct} \tag{13}$$

$$T_{pcm} = \frac{\overline{f_4(t)}}{d}(1-e^{-dt}) + T_{pcmdi}e^{-dt} \tag{14}$$

where

$$c = \frac{\left[\frac{h_{bw}h_b^{'}}{h_b^{'}+h_{bw}}\right] + \left[\frac{h_{1w}h_o}{(h_{1w}+h_o)}\right]}{M_w C_w}A_b$$

$$d = \frac{(h_b^{'}+U_{ins})}{M_{equ}}A_b$$

$$f_3(t) = \left(\frac{h_{1w}h_o}{(h_{1w}+h_o)}\right)T_a A_b + \frac{\left[h_b^{'}h_{bw}T_{pcm}A_b\right]}{(h_{bw}+h_b^{'})}$$

$$f_4(t) = \left[h_b^{'}T_b + U_{ins}T_a\right]A_b$$

The detailed procedure for energy balance equations for the system without PCM can be followed from [93] to estimate temperatures, yield, and efficiency. The hourly yield per unit area may be estimated as;

$$\dot{m}_{ew} = \frac{h_{ew}\,(T_w - T_{gi}) \times 3600}{L} \tag{15}$$

4. Performance augmentation using PCMs

This section encompasses a broad range of subject areas, highlighting the significant diversity of PCMs and solar still configurations explored

in the literature. The results of incorporating PCMs on various configurations of solar still are analyzed, and findings are presented to provide insights using various energy storage materials, including combination with sensible heat storage/dispersed nanomaterials.

4.1. Passive solar stills

To estimate the effect of stearic acid on the performance of a solar still, El-Sebaii et al. [58] formulated heat balance equations for a solar still using stearic acid in a crevice beneath the basin liner and numerically solved. The basin was 1 m^2, fabricated using a black painted galvanized iron sheet of 0.002 m thickness, leaving a crevice beneath for PCM. The bottom and sides of the solar still were insulated using sawdust layer of 0.03 m thickness housed in a 0.01 m thick wooden frame. The condensing cover was made of 0.003 m thick window glass, inclined at an angle of 22.76°, equal to the latitude of the place. An increase in PCM mass has significantly improved overnight and daily yield. Moreover, the daytime yield has decreased due to the enhanced heat storage during sunny periods. For 0.033 m depth (30 kg) of PCM beneath the basin liner, the evaporative heat transfer coefficient was increased by 27 %, with an enhanced daily yield of ~9.0 kg/m^2 and thermal efficiency of 85.3 %. The increase in yield was 80.1 % higher than the solar still operating without PCM (~5.0 kg/m^2). However, PCMs were found to be more effective for low water mass during the winter, further observing that for a deep basin solar still (water mass > 100 kg), the incorporation of 15 kg of stearic acid yields maximum during the night due to considerable storage of sensible heat in basin water during daytime. To explore the irreversibility during PCM's storage/discharge period, Asbik et al. [95] conducted an exergetic assessment of single slope solar still. Using paraffin wax (PCM), higher irreversibility occurred during daytime compared to night, which is crucial in the basin liner and PCM during the energy storage phase, and its variation depends on PCM depth beneath the basin liner. Maximum exergetic efficiency can reach >30 % by varying PCM depth (0.005 to 0.05 m) and wind velocity, excluding the heat storage period. The exergy efficiency of passive solar still was found to be comparatively lower using PCMs because a large part of latent heat remains unutilized during discharging mode. The effect of PW mass of 18 kg (0.02 m thick) on weir-type cascade solar still was experimentally investigated [96]. The absorber of solar still consisted of 15 steps (0.6 × 0.05 × 0.03 m dimension) with the absorber and evaporation surface of 0.72 and 0.45 m^2, respectively. The absorber plate was painted black with a weir of 0.02 m height and placed into an aluminum frame with a glass cover tilt of 30°. The feed water was injected continuously with a 0.07 kg/min flow rate. They found a higher daily yield of 6.7 kg/m^2, which was ~31 % higher than the yield obtained from a system without PCM, which yielded 5.1 kg/m^2. The overall energy efficiency of the still with PCM and without PCM was estimated as 64 % and 47 %, respectively.

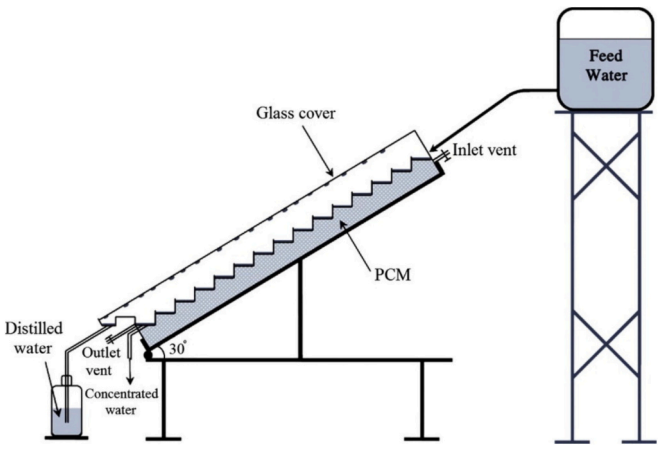


Fig. 6. Diagrammatic representation of weir type solar still [94]. Adapted with permission from Elsevier.

Further, Kumar et al. [97] tested weir-type cascade solar still of 0.20 m^2 basin area incorporating PCMs, namely bee wax, stearic acid, and palmitic acid, having melting points of 61.0, 55.0, and 54.3 °C, respectively. Area of each weir was 0.01 m^2, glass tilt of 30°, and flow rate of inlet water was maintained as 2.9 ml/s. The yield and energy efficiency of the solar still incorporated with palmitic acid has been found to be 2.35 kg/m^2 and 24.58 %, respectively, due to its low melting point of palmitic acid, and found higher by 33.69 % and 35.46 %, respectively, than the solar still without PCM. Energy efficiencies using bee wax and stearic acid were 21.7 and 20.6 %, respectively. In continuation, a theoretical assessment of the same solar still using 18 kg PW under sunny and semi-cloudy was carried out for the system, as shown in Fig. 6 [94]. The solar still incorporated with PCM yielded 4.94 kg/m^2 with energy and exergy efficiency as 76.69 % and 8.59 %, respectively. While yield, energy and exergy efficiency from solar still were found to be 3.84 kg/m^2, 74.35 %, and 6.53 %, respectively, without using PCM. Solar still without PCM yields higher during daytime than still with PCM because absorbed solar energy is utilized to increase PCM temperature, leading to low evaporative heat transfer between water and glass cover. The PW-loaded system performed better during partial cloudy conditions because evaporative exergy in the still without PCM is lower than still with PCM during low solar radiation intensity. The absorber plate has the highest irreversibility rate, accounting for ~78.8 and 83.1 % of the total irreversibility for the system with PW (semi-cloudy) and without PW (sunny day), respectively.

Pyramid-shaped solar still yields higher than conventional solar still. Keeping in view, the effect of using PCM (PW) on a GI sheet (0.0015 m thick) fabricated pyramid-shaped solar still of 0.64 m^2 basin area with side height 0.15 m and a v-grooved absorber plate was studied [98]. The side walls were insulated with fiberglass (0.05 m), which was placed within the wooden box of 0.03 m thickness. The PCM (44 kg) was filled between a square basin made from 0.001 m thick copper material and a v-grooved absorber. The glass cover of 0.003 m thickness was tilted at 30.47°. For 0.02 m basin water depth, the total yield obtained from the PCM-loaded solar still was 6.6 kg/m^2, higher by 87.4 % than the yield obtained from the pyramid-shaped solar still without PCM (3.5 kg/m^2). Furthermore, Kabeel et al. [61] theoretically examined the influence of various PCMs such as stearic acid, capric-palmitic acid, calcium chloride hexahydrate, A-48, etc. on conventional SS. The results revealed an increase in yield by 92 % and 100 % by incorporating 0.04 m thick organic A 48 and inorganic capric-palmitic PCM, respectively, as solar still without PCM. In addition, no remarkable effect on productivity has been observed with PCM thickness, so a small thickness is recommended to reduce the cost. The capric-palmatic acid has a low melting temperature of 22.5 °C, but its low thermal conductivity with high latent and specific heats enables it to give high storage energy compared to A48. The payback estimated from the solar still incorporated with various PCMs is depicted in Fig. 7. However, A-48 was recommended because of its eco-friendly nature and minimal effect of PCM thickness on the performance. A 43.3 % increase in yield was reported using PW-stabilized material on pyramid solar stills [44].

To investigate the results of basin materials and PCM, Vigneswaran et al. [99] used two materials to fabricate single slope solar still and study the effect of bottom and side losses, i.e., still made of galvanized iron (TSS), incorporated with PW (TPSS) and another made of acrylic with PW (APSS). The PW was filled in the annulus of 21 GI tubes (0.286 kg in each tube) and placed over the basin liner in basin water. It was inferred that the usage of acrylic as basin material reduced the charging time of the PCM but also delayed the discharging time of the PCM, resulting in enhanced yield. The daily yields of 4.36, 3.9, and 3.66 kg/m^2 with exergy efficiencies of 3.56, 2.99, and 3.46 % were obtained from APSS, TPSS, and TSS systems, respectively. The production cost of distillate from APSS (Rs. 1.23) was higher than the distillate cost produced by TSS (Rs. 0.67) and TPSS (Rs.1.09) due to higher fabrication

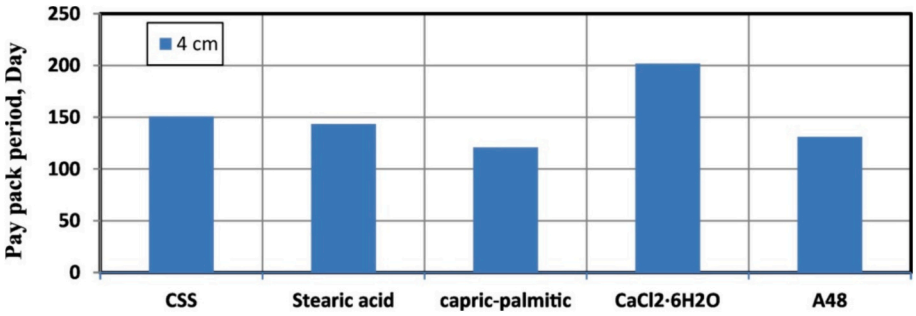


Fig. 7. Payback time using various PCMs in solar still [61]. Adapted with permission from Elsevier.

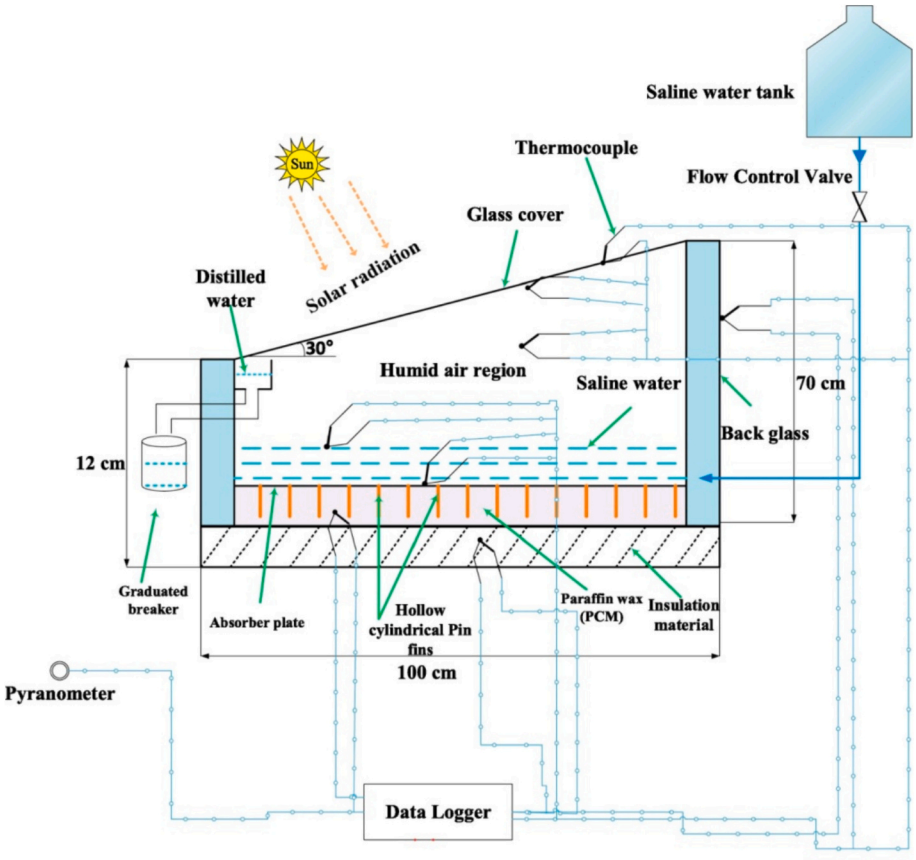


Fig. 8. Schematic of PCM with fin configuration in solar still [100]. Adapted with permission from Elsevier.

costs using acrylic. It was recommended that wicking material be placed on the water surface to overcome the low yield from TPSS and APSS during sunshine hours compared to TSS.

To study the combined effect of latent and sensible heat storage materials, the performance of conventional solar still was investigated using various arrangements such as (a) conventional (case-1), (b) PCM loaded (17 kg paraffin) (case-2), (c) with hollow cylindrical pin fins attached to PCM (case −3), and (d) PCM along with steel wool fiber in the basin (case-4), as shown in Fig. 8 [100]. The basin liner thickness of 0.004 m is painted black to capture maximum radiation, and armaflex insulation prevents heat dissipation from the basin liner to the ambient. The top cover and lateral walls were fabricated from a transparent glass of 0.005 m thick. PCM's reservoir made of a GI sheet of 0.026 m was placed beneath the basin liner and employed with 17 kg of PW to enlarge the heat exchange contact with the liner. The hollow cylindrical pin fins occupy a smaller volume compared to PCM and also enhance the heat transfer between the fins and PCM by maximizing the projected surface area, resulting in augmentation of the rate of heat storage and retrieving energy/exergy transfer from and to the saline water. The system under case-3 yielded higher by 17.0 % and 7.0 % compared to the solar still under case-1 and case-2, respectively. Solar still under case-2 has enhanced the yield by 7 % during the day and 400 % during the night, compared to the solar still under case-1. The solar still under case-4, has enhanced the daytime yield by 14 % compared to the case-2. This is due to the enhanced heat transfer within basin water, which is attributed to increased internal convective and evaporative heat transfer coefficients. However, in this arrangement, less energy is available during the day for the PCM to store, resulting in a decrease in night yield by 80 % compared to case −2. The production cost of distillate was found to be 0.427, 0.51, 0.54, and 0.50 $/kg for cases 1 to 4, respectively [Fig. 9]. The energy efficiency of solar still under case-2, 3, 4, and 5 enhanced by 9.4, 18, 16 and 28 % compared to case-1, while the exergy efficiency using case 4 increased by 10, 50, and 68 % compared to case-1, 2, and 3, respectively. Placing steel wool fiber in the basin significantly enhances the daytime

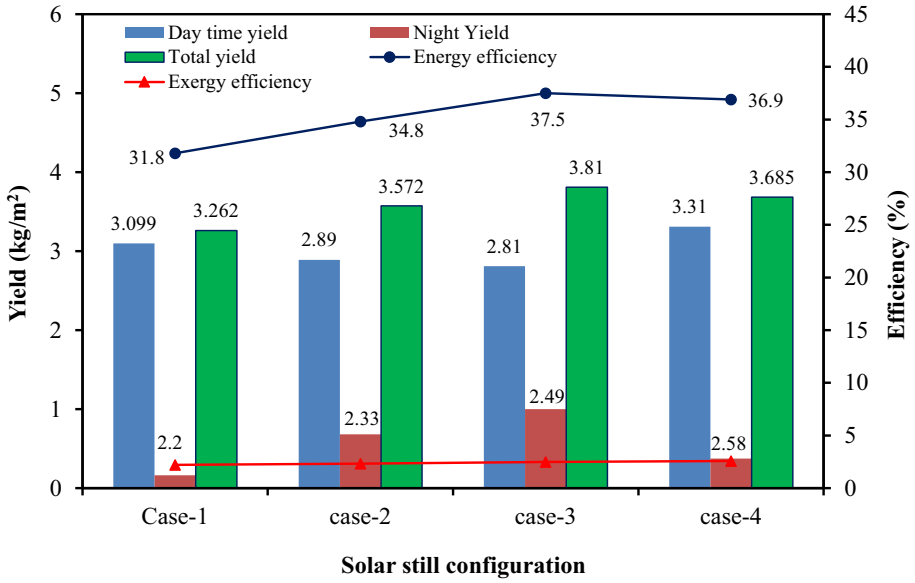


Fig. 9. Comparative overall performance of PCM and sensible heat storage material on solar still.

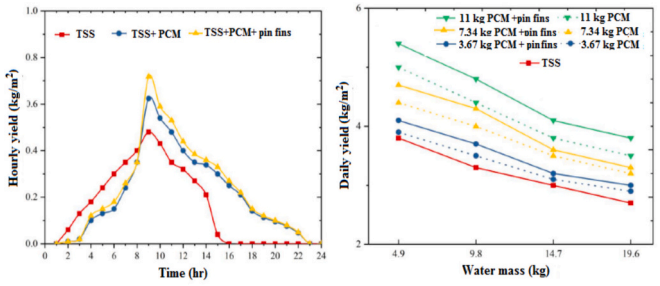


Fig. 10. Effect of palmitic acid and basin water on the productivity [101]. Adapted with permission from Elsevier.

energy and exergy efficiencies, however, considerably reducing these values at night. The steel wool fiber compensates for the evaporation loss during the daytime (due to heat transfer from basin to PCM) owing to the increased effective convective and radiative energy transfer area.

Further, Kateshia et al. [101] investigated the effects of different organic fatty acids such as lauric acid, palmitic acid and stearic acid along with high-conductive aluminum pin fins which were inserted in the PCM storage to enhance the rate of heat conduction. The solar stills were fabricated with a 0.002 m thick GI sheet of 0.49 m^2, painted black, and insulated using 0.005 m thick glass wool. A 0.004 m thick glass cover, tilted at 23° was placed as a condensing cover. The yields from the solar still with pin fins embedded PCM and solar still with PCM were higher by 30 % and 24 %, respectively, compared to solar still without PCM and pin fins. The yields obtained during low/off sunshine hours are depicted in Figs. 10–11. The yield, energy efficiency, and exergy efficiency were observed maximum as 4.43 kg/m^2, 47.9 % and 2.23 % using lauric acid and pin fins in winter. While maximum using stearic acid and pin fins in summer as 5.74 k/m^2, 56.2 % and 2.96 % with production cost 0.016 $/L and 72 days payback time, respectively. The application of Al_2O_3 dispersed water and PW underneath the absorber has also been reported using SS of 1 m^2 of absorber plate made of 0.0015 m aluminum sheet, insulated by 0.002 m thick plywood and 0.004 m thick glass cover inclined at 26° over the still [102]. Al_2O_3 dispersed water with PCM enhanced the daily from 5.1 to 5.47 kg/m^2 compared to PCM loaded solar still alone. Higher yield using Al_2O_3 dispersed water and PCM system is due to an increase in heat transfer coefficient with the basin water during sunshine hours and higher water temperature during night

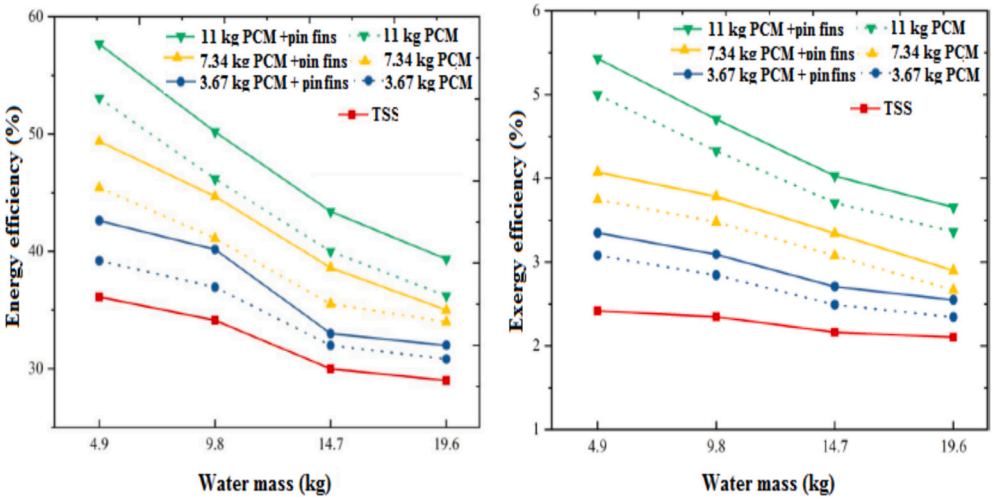


Fig. 11. Effect of palmitic acid and water mass on energy and exergy efficiency [101]. Adapted with permission from Elsevier.

(during discharge of PCM), resulting in the increased convective and evaporative heat transfer coefficients from water to glass cover. Furthermore, the effect of micro-flakes of graphite and CuO with different concentrations (0.125–2 %), different water depths (0.0025–0.05 m), and film cooling flow rates (1 and 12 kg/h) investigated by Sharshir et al. [103]. The still was made of 0.0015 m thick iron sheet of 0.25 m^2 basin area and insulated using 0.05 m thick fiberglass. The thickness of glass cover was 0.0035 m and tilted 30°, equal to the latitude of the place. The yield enhancement by 53.95 % and 44.19 % using graphite and CuO micro-flakes, respectively, which enhanced further to 57.6 % and 47.8 %, respectively, with glass cover cooling. Based on the particle size, 1 wt% has been obtained as the optimal concentration, beyond which no marked benefit was obtained. At the water nanofluid depth of 0.0025 m, the lower yield is obtained due to the creation of dry spots, while yield increases for 0.005 to 0.01 m water depth due to faster evaporation and condensation. The study was further extended using paraffin wax, mixing the graphite flakes with basin water and with glass cover cooling, and with basin water, and the yield improved by 73.8 % compared with the conventional still [104].

To test the effect of PCM and nanofluid combination, a hemispherical solar still was fabricated using 0.001 m thick galvanized steel of 0.34 m diameter and 0.04 m height, placed in a circular wooden container of 0.38 m diameter and coated with black silicone rubber, maintaining 0.02 m gap between bottom and side for PCM [47]. A 0.003 m thick hemispherical plastic cover was placed over the basin. Four stills, namely conventional hemisphere, hemisphere with PW (2 kg), hemisphere with 0.3 wt% CuO/water nanofluid, and hemisphere with PW + 0.3 wt% CuO/water nanofluid were tested, as shown in Fig. 12. Maintaining basin water depth of 0.01 m, the yield has been enhanced by 60.41 % and 29.17 % from the solar still loaded with nanofluid and PCM, respectively, compared to conventional hemisphere solar still. A hybrid combination of PCM and CuO/water nanofluid further enhanced the yield by 80.20 % and reduced the production cost by 75 %. The energy efficiency of respective hemisphere solar stills has been estimated as 35.52 %, 45.45 %, 56.46 %, and 63.61 %, respectively.

The enhancement of yield by employing various PCMs on various configurations of passive design of solar stills compared to solar still without PCM under different geographical locations is shown in Fig. 13. A considerable enhancement in yield has been found using PCMs, depending on their selection, position, and system configuration. An increased yield in the order of capric-palmitic (100 %) > A-48 (92 %) > stearic acid (68 %) was reported [61]. The capric-palmitic has high latent and specific heats, low melting point (22.5 °C), and improves energy storage, response time, and heat transfer, with reduced heat losses compared to other PCMs. However, A-48 PCM is preferred due to its less environmental impact and cost less compared to capric-palmitic. Using stearic acid as a PCM, the productivity has been enhanced by ~80 % compared to the CSS without PCM. While PW enhanced yield by ~87 %, using pyramid-shaped solar still has been reported, possibly due to the high latent heat of fusion and thermal conductivity of PW compared to stearic acid. However, the choice between stearic acid and PW depends on the desired temperature range, thermal conductivity, cost, and availability. The study has further revealed that most of the studies used organic PCMs such as paraffin wax, and more work is needed on other types of PCM in the specific field. Fabricating materials also play an essential role in enhancing the performance of solar stills. The acrylic sheet made solar still with PW yielded higher by ~11.0 % compared to the solar still made of GI sheet with PW, possibly due to the transfer of more radiations to the basin water [99]. The variations in the performance using the same PCM for similar solar stills may be due to the variable operating and climatic parameters, including designs of solar stills tested. The hybrid operations of solar still using PCM and nanofluids/steel weir mesh have been found beneficial. The PCM acts as a heat source during night to augment the night yield, while CuO nanoparticles work as a thermal enhancer for harvesting solar energy by the saline water due to the higher thermodynamic stability and thermal conductivity.

4.2. Active solar stills loaded with PCMs

To enhance the productivity, solar stills are fed by preheated water to operate at high temperatures. To investigate the effect of PCM on high-temperature desalination, Arunkumar et al. [111] incorporated 1.2 mm thick black-painted paraffin wax-filled copper balls in a hemispherical solar still integrated with a concentrator (Fig. 14). On a typical sunny day, the system loaded with PCM yields 4.46 kg/m^2, about 26 % higher than the system operating without PCM (3.52 kg/m^2). The metallic balls act as sensible heat storage for the increase in yield during the day.

The impact of PCM on solar still coupled with oil serpentine loop from a cylindrical parabolic concentrator and loaded with PW beneath the basin was investigated by Kabeel and Abdelgaied [112]. A CSS fabricated using a 0.0015 m thick GI sheet of basin area 0.72 m^2 was insulated with fiberglass (0.05 m thick). A commercial glass cover, with a thickness of 0.003 m, was positioned at the top of the body at 30.47^o. The absorber vessel made from a copper sheet (0.0004 m thick) was placed inside the basin, and the gap was filled with a PCM of 0.03 m thickness. The oil heat exchanger consists of a black-coated single copper tube with inside diameter of 0.0125 m, 12 passes 8.4 m long, and immersed in the basin water. The concentrator has a 1.2 m wide and 2 m long dimension, an aluminum reflector with a receiver of copper tube 2 m long and 0.019 m diameter covered by an evacuated tube of 0.038 m diameter. The enhanced daily yield from ~4.48 to 10.77 kg/m^2 was obtained using the modification [Fig. 15]. The yield enhanced by ~140.4 % with a decreased production cost from 0.0177 to 0.0174 $/kg

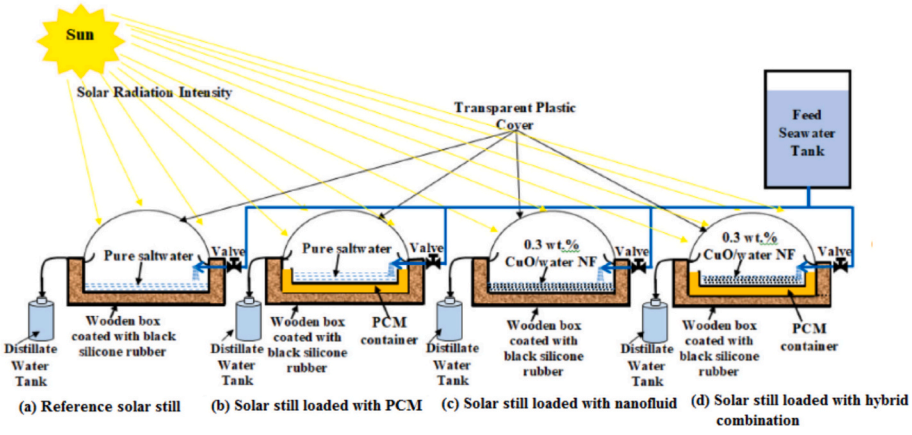


Fig. 12. Schematic of hemispherical solar stills [47]. Adapted with permission from Elsevier.

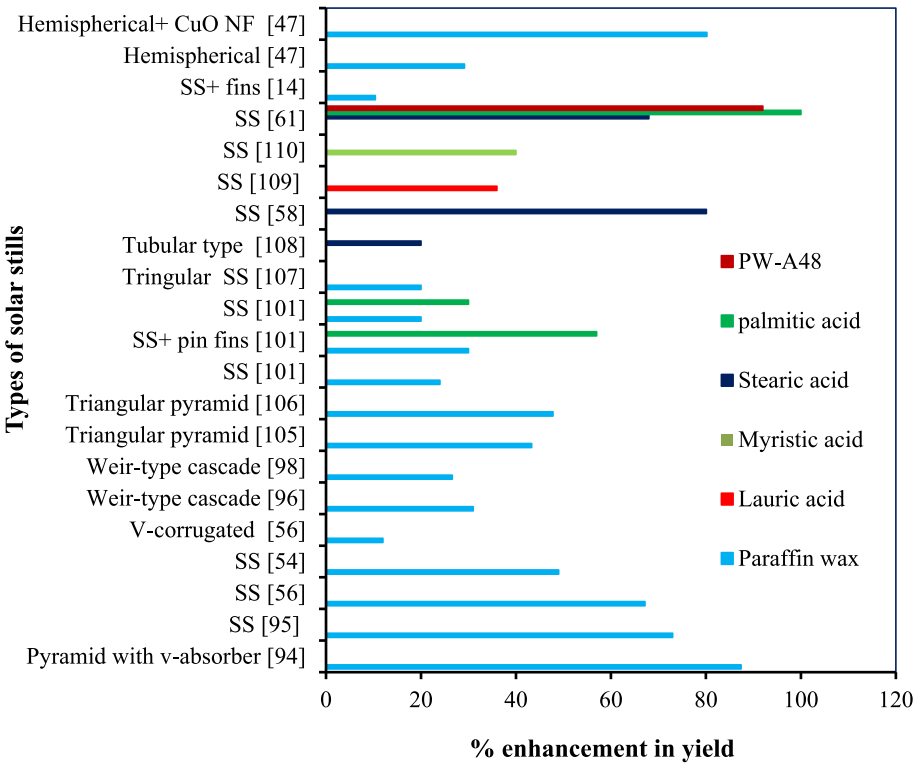


Fig. 13. Comparative enhancement of yield using PCMs on various solar stills [14,47,54,56,58,61,94–96,98,101,105–110].

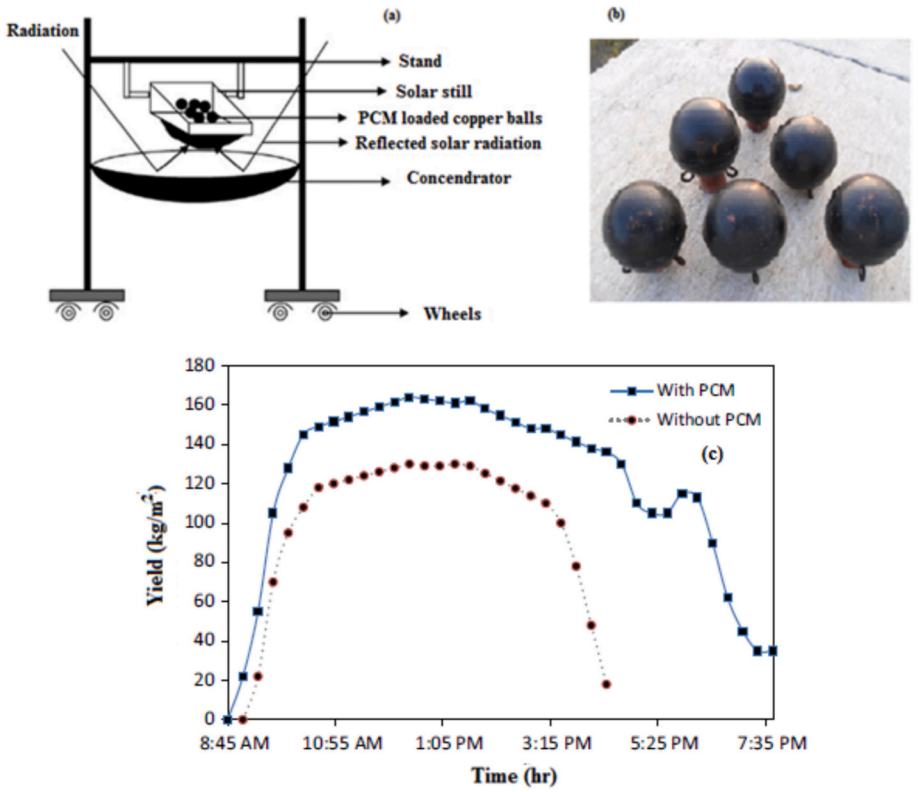


Fig. 14. (a) Schematic of concentrator coupled solar still, (b) PW loaded balls, and (c) variation of yield with and without PW [111]. Adapted with permission from Elsevier.

than the conventional solar still. However, the overall daily efficiency was found to be reduced by ~44 % due to the increase in solar collection area.

Further, Faegh and Shafii [113] incorporated two different mechanisms for the system consisting of two evacuated tube collectors, a saline water basin, a tank and paraffin wax, a fan, and also twenty

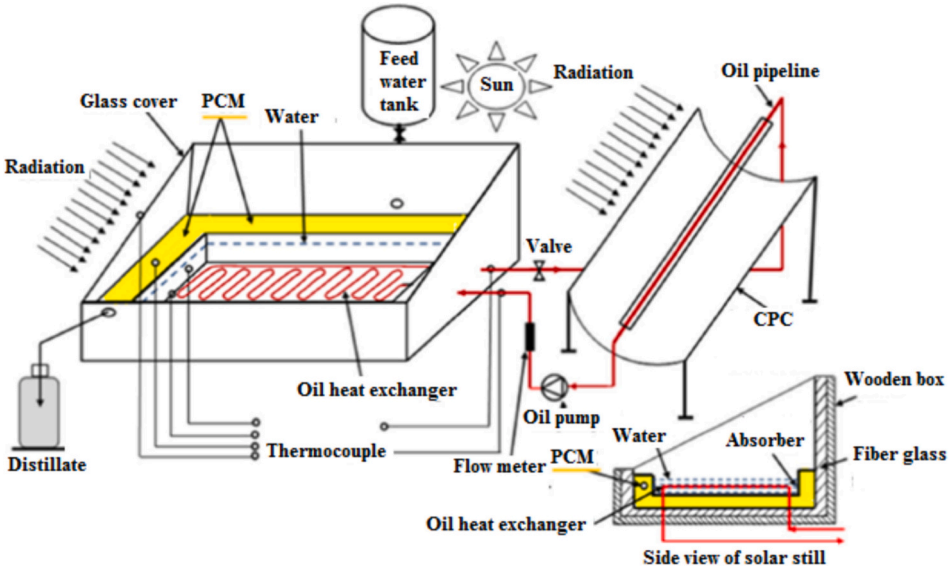


Fig. 15. Solar still integrated with CPC and loaded with PW [112]. Adapted with permission from Elsevier.

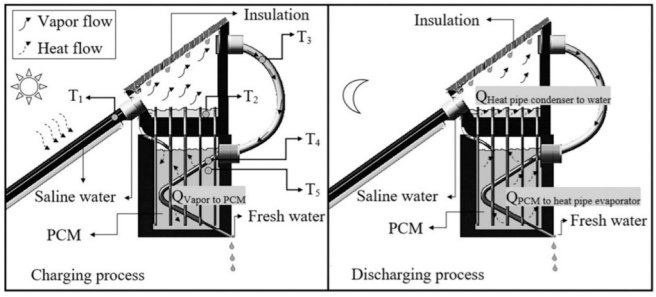


Fig. 16. Diagram of ETC integrated solar still with PCM loaded external condenser [113]. Adapted with permission from Elsevier.

thermosiphon heat pipes as depicted in Fig. 16. A fan conducts the vapor from the basin through the PCM by suitable connecting tube, insulated by polyurethane of thickness 0.03 m. PW-charged the external condenser and discharged heat to heat pipes. The evaporative section of the heat pipes lies inside the PCM, and its condenser section was placed inside the basin water. The daily yields of 3.53, 3.73, 5.49, and 6.55 kg/m^2 were obtained from the respective case as shown in Fig. 17. Daily yield increases by 86 % from the system modified under case 4. Energy efficiency of 24.0 %, 25.43 %, 41.86 %, and 50.0 % have been estimated for cases 1 to 4, respectively, with the distillate production costs of 0.021, 0.027, 0.138, and 0.015 $/kg from the respective system. From the results, it can be concluded that PCM incorporated solar stills yield more effectively once incorporated with other external modifications such as condensers. In another arrangement, ETC-integrated solar still incorporated with PW and found maximum yield, energy efficiency, and

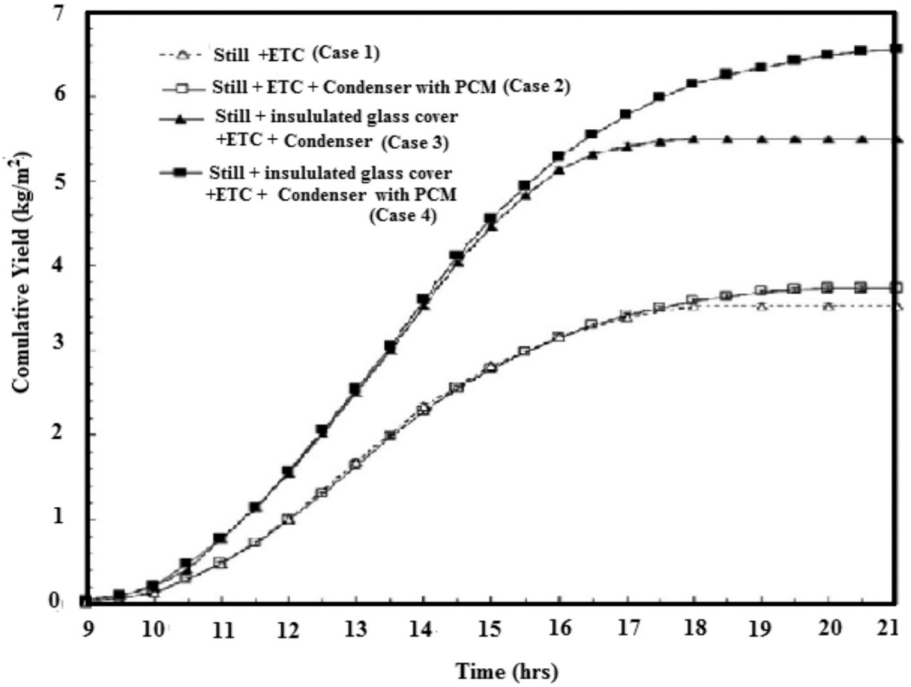


Fig. 17. Cumulative yields in various combinations [113]. Adapted with permission from Elsevier.

distillate production as 6.55 kg/m^2 with 50 % energy efficiency and a distillate production rate of 0.015 $/kg.

Al-harahsheh et al. [114] investigated the effect of inorganic sodium thiosulfate pentahydrate (STP) on active solar still. A solar still of 1 m^2 basin area has a double glass cover at 0.01 m apart. The cooling of the lower glass cover increases condensation due to enhanced driving force. A heat exchanger consisting of an aluminum coil of 0.01 m diameter was connected to a flat plate collector of 2.5 m^2 consisting of 10 copper tubes of 0.008 m diameter. A stainless steel box tub contains 15 plastic tubes and attached to the bottom of the solar still and filled with PCM. This inorganic sodium thiosulfate pentahydrate (STP) is considered due to its high latent heat of fusion, small volume change upon melting, ease of availability, and low cost compared to other competitive PCMs. A significant enhancement in yield during the night (40 % of accumulated yield) was obtained. The increase in flow rate through the collector from 2 to 30 ml/s enhanced the yield three times and found to be more effective at lower water mass in the basin due to the increased heat removal rate by the water flowing in the FPC tubes with an increased flow rate. The influence of inorganic PCM (sodium thiosulfate pentahydrate) to water ratio on the performance of glass cover cooled solar still integrated with PV/T flat plat collector has been studied as depicted in Fig. 18. The increase in PCM to water ratio from 10 to 100 % decreases the yield by ~30 % due to more storage of PCM energy, resulting in a decrease in the basin water temperature during solar time and less daytime productivity. The increased flow rate of cooling water from 0.01 to 0.1 kg/s increased the yield by 37 % due to an increase in the temperature difference between water and glass cover, resulting in an increase in the evaporative heat transfer coefficient. Furthermore, at low depths of basin water, PCMs have been more effective, and yield was doubled by decreasing the depth from 0.1 to 0.05 m. Mehdiabadi et al. [115] studied the double slope solar still integrated with a single PV/T collector and incorporated with PW. Increased yield and exergy efficiency by 10.6 % and 27 %, respectively, were found with an increase in the water circulation rate (0.001 to 0.01 kg/s) during summer. As shown in Fig. 19, the exergetic efficiency increases linearly with the increase of PCM and basin water mass. Furthermore, the effect of the water/PCM ratio has been investigated using PW, stearic acid, and lauric acid filled in the annulus of 6 copper cylinders placed in basin water [116]. The drop in water temperature by 9.2, 17.6, and 21.5 % have been reported using PW, stearic acid, and lauric acid, respectively, with the rise in water depth from 0.01 to 0.05 m with a maximum increase yield 1.2 kg/

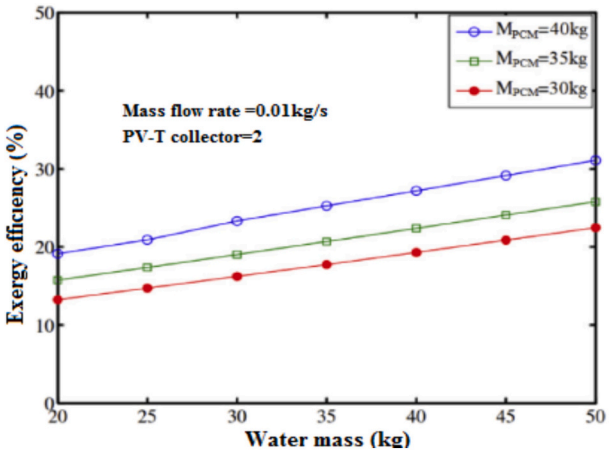


Fig. 19. Impact of water mass and PCM on exergetic efficiency of PV-T integrated dual slope still [115]. Adapted with permission from Elsevier.

m^2 due to paraffin wax, flowed by 1.0152 kg/m^2 (stearic acid) and 0.932 kg/m^2 due to lauric acid. This is due to the lower melting point of lauric acid (42–46 °C) as compared to paraffin wax (58–60 °C) and stearic acid (52–56 °C).

Mazraeh et al. [117] integrated single slope solar still with PV module, evacuated tube collectors, and PCMs (PW, Gallium, PWRT44HC) to evaluate the effect of PCM-PV modules, water depth, and number of collector tubes on the performance. ETC collector with an efficiency of 0.536 consists of concentric tubes with inner and outer diameters of 0.044 m and 0.047 m, respectively, and a length of 1.4 m and operates in natural mode as shown in Fig. 20. The incorporation of PCM has improved the thermal efficiency with negligible effect on exergy efficiency. The use of PW for 0.03 m water depth and 30 ETC tubes yields 4.5 kg/m^2, significantly higher (20.32 %) than the system without PCM (2.3 kg/m^2). The maximum energy and exergy efficiencies were estimated at 17.93 % and 6.95 %, respectively, using PW with ten tubes. When the PV was placed, only fractions of radiation entered the basin directly, which had a negligible effect on increased heat storage within the PCM during sunshine hours.

Abed et al. [118] have extended the investigation with various combinations such as conventional design (case-1), conventional with

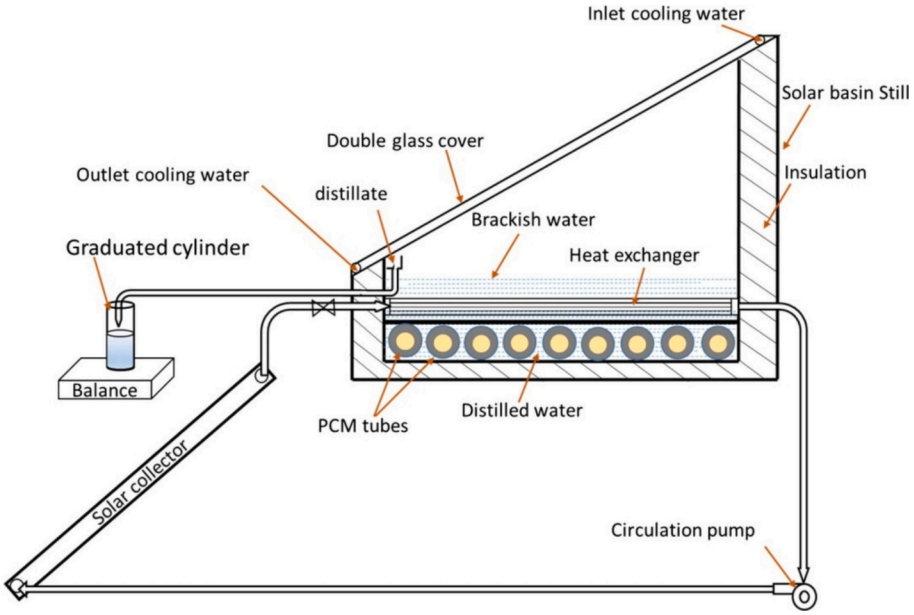


Fig. 18. FPC coupled solar still incorporated with PCM and glass cooling [114]. Adapted with permission from Elsevier.

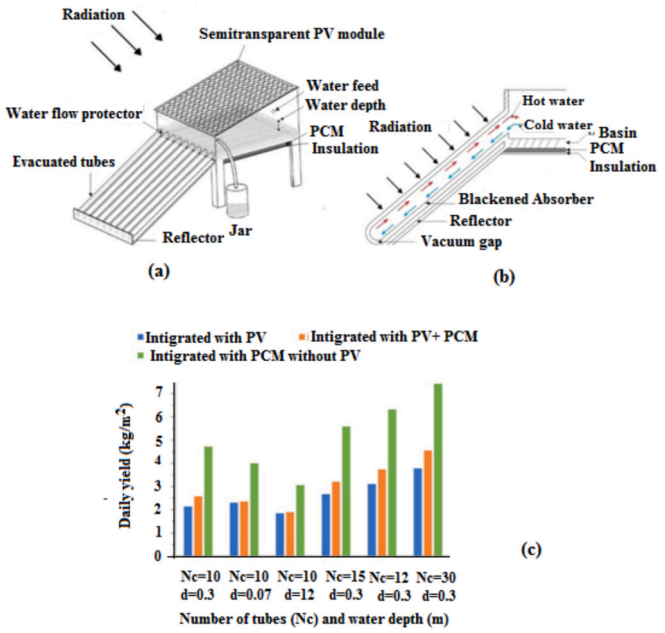


Fig. 20. Schematic of (a) solar still (b) evacuated tube, and (c) PCM effect [117]. Adapted with permission from Elsevier.

cylindrical paraffin capsules (case-2), conventional with cylindrical paraffin capsules and coupled with the collector (case-3), conventional with cylindrical paraffin capsules and coupled with collector and UV vaporizer (case-4), with the arrangement shown in Fig. 21. The rectangular solar still was fabricated with a wooden area of 0.5 m^2 and 0.004 m thick glass cover fixed at an angle of 35° on the top. The basin was fabricated from 0.0015 m thick GS, and the bottom and side walls were insulated externally using plywood with a thickness of 0.01 m. 20 PW cylindrical capsules of 0.025 m diameter and 1 m length were positioned inside the basin water. The solar collector was a rectangle lope in which the water flowed via the 1.5 m tube. On a typical summer day, the increased daily yield with PCM (case-2) was reported to be 30.6 % higher than case-1, while the solar still with case-3 yielded 244 % and 163 % higher compared to case-1 and case-2, respectively. However, including all combinations (case-4), 7.255 kg/m^2 yield was reported, which was 415 %, 340 %, and 49.9 % higher compared to cases 1, 2, and 3, respectively, with a deviation in maximum productivity time. The energy efficiency obtained in each case is shown in Fig. 22. The production cost associated with each case is estimated as 0.037, 0.056, 0.40, and 0.028$/kg, respectively, accounting for 270 clear days.

The stepped solar still fitted with an external condenser was fabricated to study the impact of PCM and condenser on yield [15]. The modified solar still consisted of 7 steps of 0.8 × 0.08 × 0.035 m size, made of 0.0015 m thick iron sheet and 0.003 m thick glass cover inclined at an angle of 36.26^o, equal to the latitude of the place. Basin and the bottom of the steps were insulated with 0.002 m thick elastomeric material. Water vapor from the solar still was transferred to the condenser by a fan, creating negative pressure inside the comber of the solar still. The 0.001 m diameter of the copper tube with length of 2 m rolled inside the water cooled external condenser (Fig. 23). Using 2.5 kg stearic acid (PCM) beneath the steps and maintaining 0.003 m steps water, the enhancement of daily yield by 26 % (0.8 kg), 43 % (0.91 kg), and 104 % (1.3 kg) has been obtained using system incorporated with the external condenser, with PCM alone, and condenser + PCM, respectively, compared to the conventional stepped solar (0.635 kg). The results have revealed a significant contribution of PCM and external condenser on the performance of stepped solar still.

Fig. 24 represents the comparative enhancement in accumulated yield using various PCMs and active designs of solar stills, which have

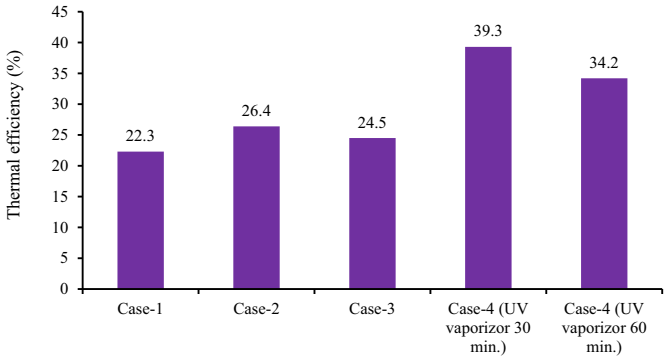


Fig. 22. Energy efficiency of solar still with various combinations, derived from [118].

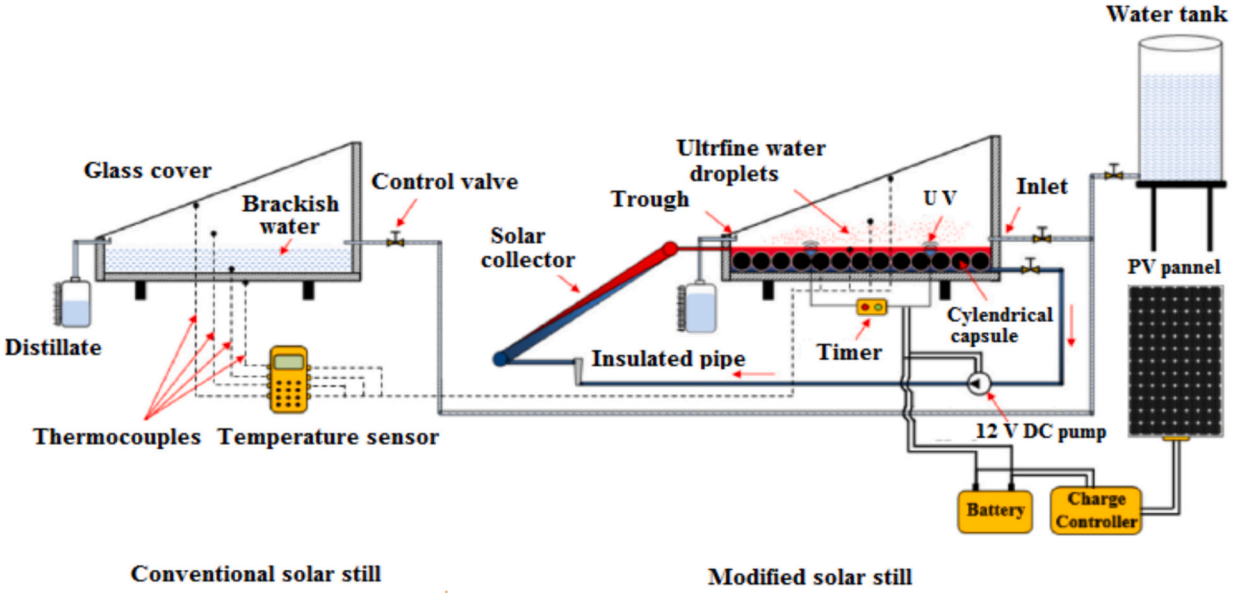


Fig. 21. Schematic of PW loaded solar still under active mode [118]. Adapted with permission from Elsevier.

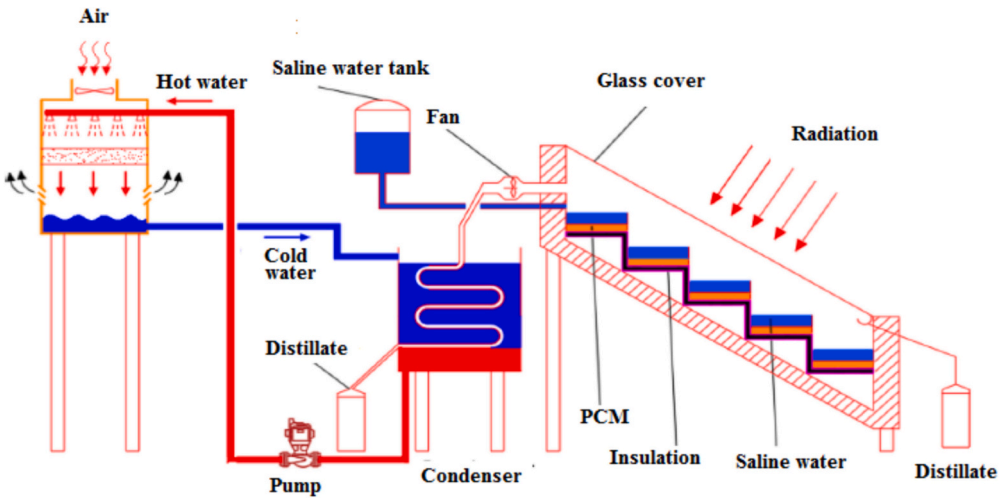


Fig. 23. Schematic of stepped solar still with PCM [15]. Adapted with permission from Elsevier.

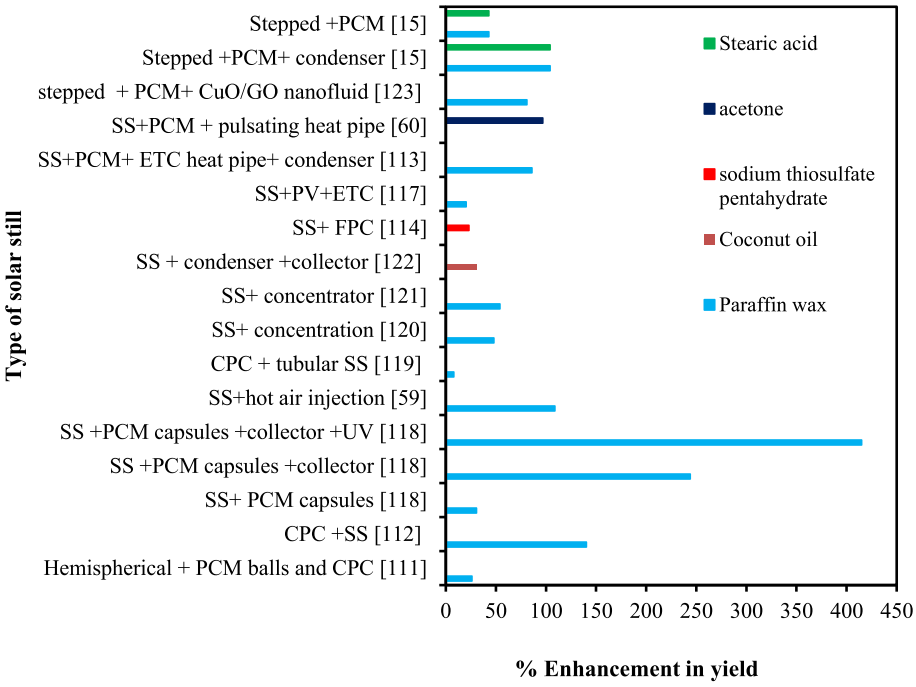


Fig. 24. Comparative enhancement in yield using PCM in active solar stills (%) [15,59,60,111–114,117–123].

been found to be more effective than passive solar stills for greater yields, subject to the type of combination employed. A significant improvement in the yield can be noticed by incorporating PCMs and collectors together, and the use of stearic acid and PW appears to be a more effective PCM for the systems researched. Incorporating PW, PW + FPC, and PW + FPC + UV systems enhanced the productivity by 30.6, 244, and 415 %, respectively, compared to the CSS. The inclusion of UV vibrator can agitate the water surface, breaking the surface tension and promoting more rapid evaporation. However, the higher operating time of the UV system adversely affects the yield due to disruption in the temperature gradient within the still, affecting the vaporization and condensation rates. Coconut oil is a green PCM and has some potential for thermal energy storage. However, its performance in solar still may be limited due to factors such as phase change temperature range, heat transfer properties, stability, etc.

5. Incorporation of NePCM in solar stills

Nanoparticles are heat carriers, improving the thermal conductivity of base fluids upon mixing. Prasad et al. [124] have achieved a maximum efficiency of 49 % from a solar still loaded with silica gel and filled with graphite nanofluid. Thermal conductivity has been reported to be improved using CNTs and CNFs compared to the use of metal or their oxides as nanoparticles alone [125]. Mehrali et al. [126] reported the enhancement in the thermal conductivity of PW to 0.9 W/mK upon mixing graphene oxide (GO). Mixing multiwall carbon nanofluid with particle sizes over 10 nm has reported a higher enhancement in the thermal conductivity of paraffin [127]. Modi et al. [128] have reported an increased yield from double slope solar still by 58.25 % and 26.6 %, maintaining 0.02 m basin water and adding 0.1 wt% of CuO and Al_2O_3 nanoparticles, respectively. Further, a decrease in water depth to 0.01 m leads to an increase in yield by 56.31 % and 26.60 %, respectively.

Al-Kayiem and Lin [129] employed PW with Cu nanoparticles (1.0

wt%) using a flat plat collector and found the system efficiency of 51.1 %, 52.0 %, and 47.6 % utilizing PCM, NePCM, and without PCM, respectively. Bait and Si-Ameur [130] provided a complete analysis of the applications of nano-enhanced solar distillation. Following [29,131], the thermophysical parameters of a micro PCM blend can be estimated as follows;

Density,

$$\rho_{nPCM} = \varnothing\rho_{np} + (1-\varnothing)\rho_{PCM} \tag{16}$$

Specific heat,

$$L_{nPCM} = \frac{(1-\varnothing)(\rho L)_{PCM}}{\rho_{nPCM}} \tag{17}$$

Latent heat,

$$L_{nPCM} = \frac{(1-\varnothing)(\rho L)_{PCM}}{\rho_{nPCM}} \tag{18}$$

Dynamic viscosity,

$$\mu_{nPCM} = 0.983e^{12.959\varnothing}\mu_{PCM} \tag{19}$$

Thermal conductivity;

$$K_{nPCM} = \frac{K_{np} + 2K_{PCM} - 2(K_{PCM} - K_{np})\varnothing}{K_{np} + 2K_{PCM} + 2(K_{PCM} - K_{np})\varnothing} + 5\times10^4\beta_k\epsilon\varphi(\rho c_p)_{PCM}\sqrt{\frac{BT}{\rho_{np}d_{np}}}f(T\times\varnothing) \tag{20}$$

where $B = 1.38\times10^{-23}$ and $\beta_k = 8.4407\times(100\varnothing)^{-1.07304}$

$$f(T\times\varnothing) = \left(2.8217\times10^{-2}\varnothing + 3.917\times10^{-3}\right)\frac{T}{T_{ref}} - \left(3.0669\times10^{-2}\varnothing + 3.91123\times10^{-3}\right)$$

5.1. Passive solar stills incorporated with NePCM

Chaichana and Kazem [132] explored the effect of nanofluid on single slope solar still and found an enhancement in yield by 10.38 % with PW, while 60.53 % using Al_2O_3-PW. This is due to the decreased melting point and increased thermal conductivity of PCM on mixing with Al_2O_3 [133]. Elfasakhany [68] fabricated 3 single slope solar stills i.e., (a) without PCM, (b) using PW (2.2 kg), and (c) PW with copper nanoparticles (40 g of 20 nm size) for the arrangement shown in Fig. 25. The rectangular basin was made of iron, insulated by a wooden plate of 0.001 m thickness with a gap of 0.02 m in between, and filled with glass wool. Another still was modified by adding a copper tube heat exchanger filled with PW, while the third still incorporated Cu/PW nano-composite in the tube heat exchanger instead of PW. The

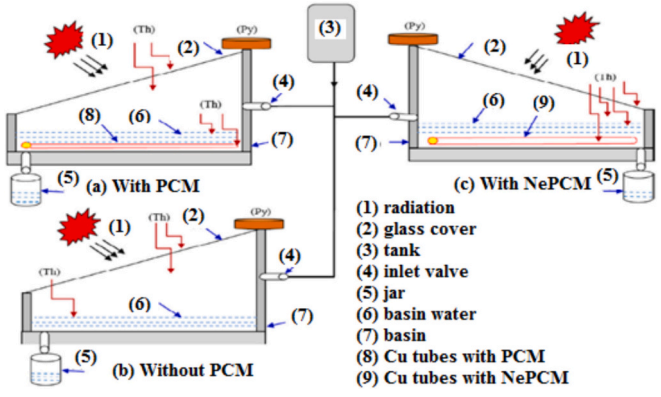


Fig. 25. A schematic of still with different concentration of nanoparticle [68]. Adapted with permission from Elsevier.

significantly enhanced yield of 119 % and 125 % with PW and PW/Cu nanofluid, respectively, has been obtained compared to still without PW. Further, during the increased radiation period, the still without PW yielded higher than those with PCM and NePCM by 9 % and 12 % due to the storage effect in PCM/NePCM. The difference in yield obtained from various arrangements has been insignificant between 12 and 14 pm, i.e., during PCM's phase transformation time.

Rufuss et al. [64] reported that solar still with NePCM (Cuo + PW) yields 35 % higher (5.28 kg/m^2) than using PW alone. Further, the effect of dispersion of various nanoparticles (TiO_2, CuO, and GO) with PW was investigated [134]. They found a significant increase in productivity by dispersing CuO, which was higher than the CSS incorporated with PW (3.92 kg/m^2), and the results are depicted in Fig. 26.

Rajasekhar and Eswaramoorthy [135] investigated the effect of Al_2O_3 mixed paraffin (NePCM) placed underneath the base plate. They observed an energy efficiency of 25, 45, and 66 % using CSS, CSS + PCM, and CSS + NePCM, respectively. The above unit was further tested for the 2 and 4 % nano concentrations [136]. The increased nano concentration has shown better energy and exergy efficiency and was found optimal at 4 % concentration with increased energy and exergy efficiency of 14 % and 11 %, respectively (Fig. 27). The combined effect of adding hybrid material with PCMs has been examined by [137,138]. The yield was reported to be enhanced by 37.55 %, while energy and exergy efficiency increased by ~37 % by using PW+ carbon gravel compared to the solar still with an inside reflector. The increase in yield using hybrid material has been investigated as 8.52 kg/m^2 by mixing 20 % graphite concentration, which was significantly higher than the yield obtained from CSS (4.38 kg/m^2). The yield and efficiency were also improved by 94.52 and 65 %, respectively, significantly higher than the unmodified stills. The efficiency of modified and unmodified solar still has been estimated as 62.38 % and 32.3 %, respectively. The effects of adding Cu chips and PW + CuO on a dual-slope solar still have been experimentally studied [139]. They reported an increased yield, energy, and exergy efficiency by 113, 112, and 190 %, respectively.

Chamkha et al. [140] investigated single-slope solar stills employing a blend of paraffin and carbon nanotubes (CNTs) to boost energy storage and thermal conductivity. The thermal conductivity increased by 26 % due to the CNT-doped paraffin compared to SS without CNT. The daily yields obtained from solar stills loaded with NePCM, PCM, and without PCM were found to be 5.8, 3.4, and 2.5 kg/m^2, respectively, with payback periods of 2.3, 2.7, and 3.0 years. The increase in PCM temperature with the addition of CNT nanomaterial is shown in Fig. 28 and found to be higher by 1.0–9.0 % on an average basis compared to the paraffin-loaded solar still alone.

The effect of NePCM on cascade-stepped solar still has been experimented on, as depicted in Fig. 29 [133]. The PW and GO mixture in varying concentrations (0.2 to 0.6 wt%) was encapsulated between the water tray and bottom absorber plate. The increase in yield by 25 % was obtained using the NePCM. The work was further extended for varying the shape of still's absorber plates (i.e., semicircular, triangular, and rectangular), incorporated with PW, and a mixture of PW and GO in the concentrations of 0.1, 0.3, and 0.5 wt% [141]. A better performance was observed using NePCM at 0.5 wt% GO with a semicircular basin, as illustrated in Fig. 30. The system with a semicircular basin yielded 20 % more than the rectangular basin. Moreover, the rectangular basin yielded 10 % higher than the triangular basin using NePCM.

The effect of the internal reflector, black paint, CuO nanoparticles, and CuO + PW nanofluid on tray-type solar still has been explored, as shown in Fig. 31 [142]. The enhancement in yield was found to be improved by 14.0, 57.0, 70.7, and 108.0 % in the order of CuO with paint< reflectors < reflectors + nanoparticles coating < reflectors+ nano-coating+ NePCM, compared to the conventional design. The daily yield obtained by incorporating all the modifications has been reported to be 5.0 times (~12 kg/m^2) higher than the conventional design (2.4 kg/m^2), with an energy efficiency of 51.5 %. CuO paint enhances the thermal conductivity of the basin liner, while the internal reflector

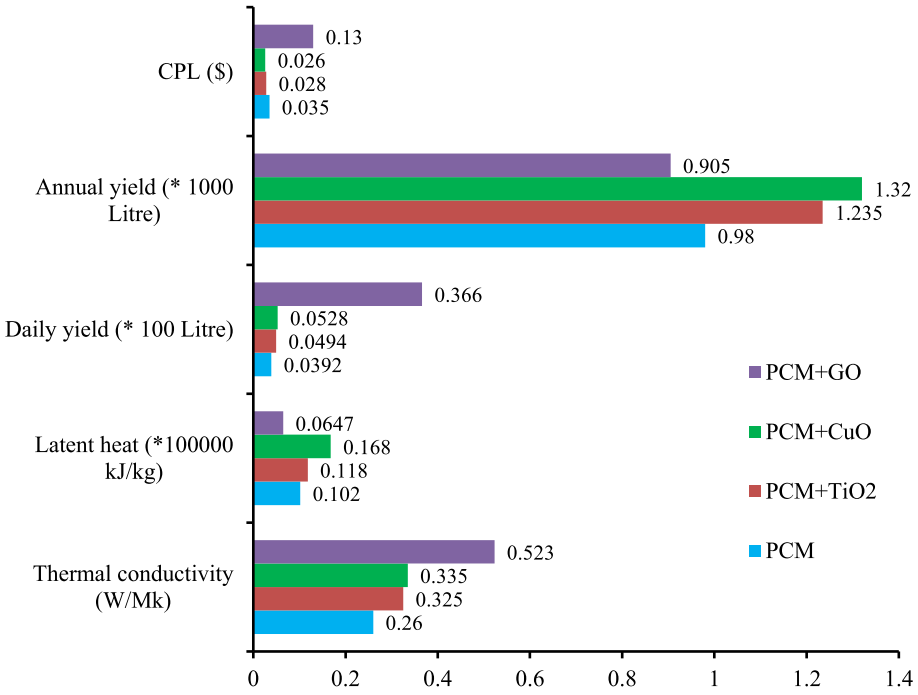


Fig. 26. Outcome of NePCM on solar still performance/m^2, derived from [134].

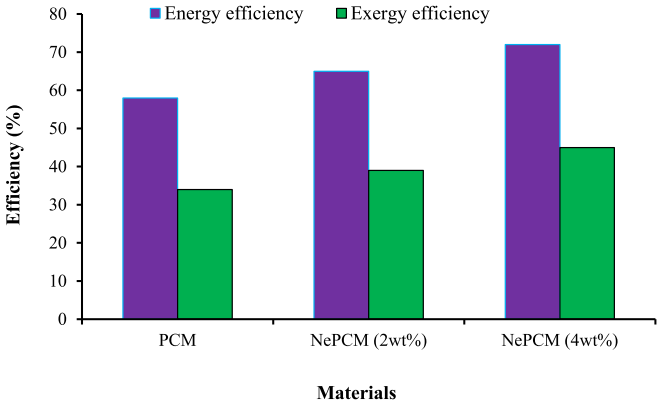


Fig. 27. Effect of Al_2O_3 concentration on energy and exergy efficiency, derived from [136].

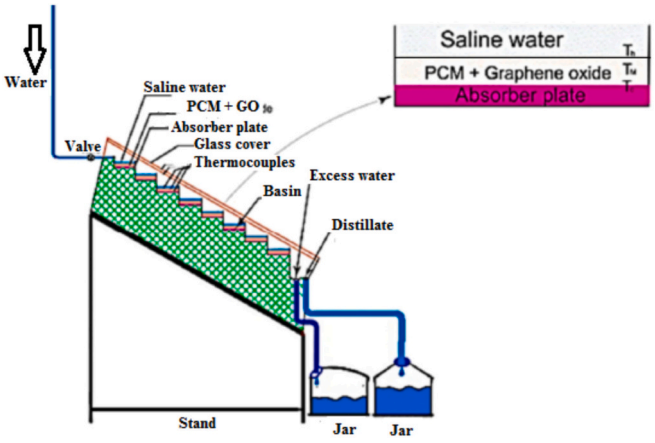


Fig. 29. NePCM loaded stepped solar still [133].

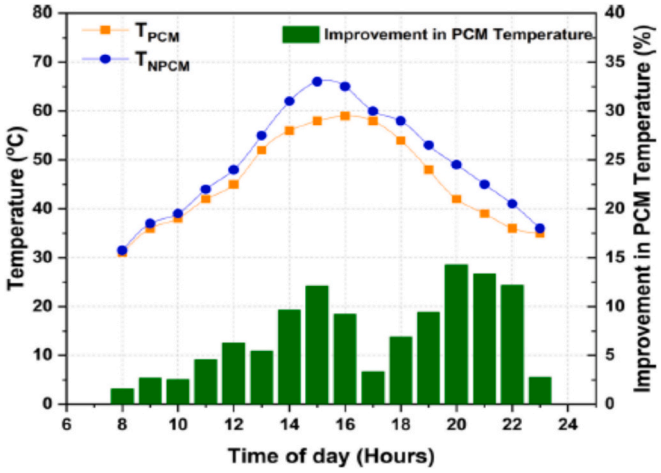


Fig. 28. Enhancement in PCM temperature by adding CNT [140].

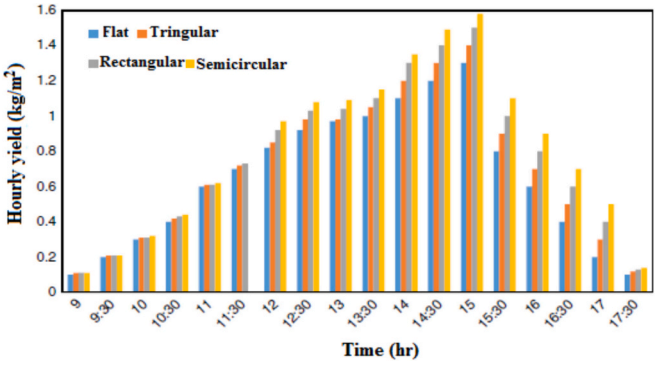


Fig. 30. Hourly yield from different shapes of absorber plate [141].

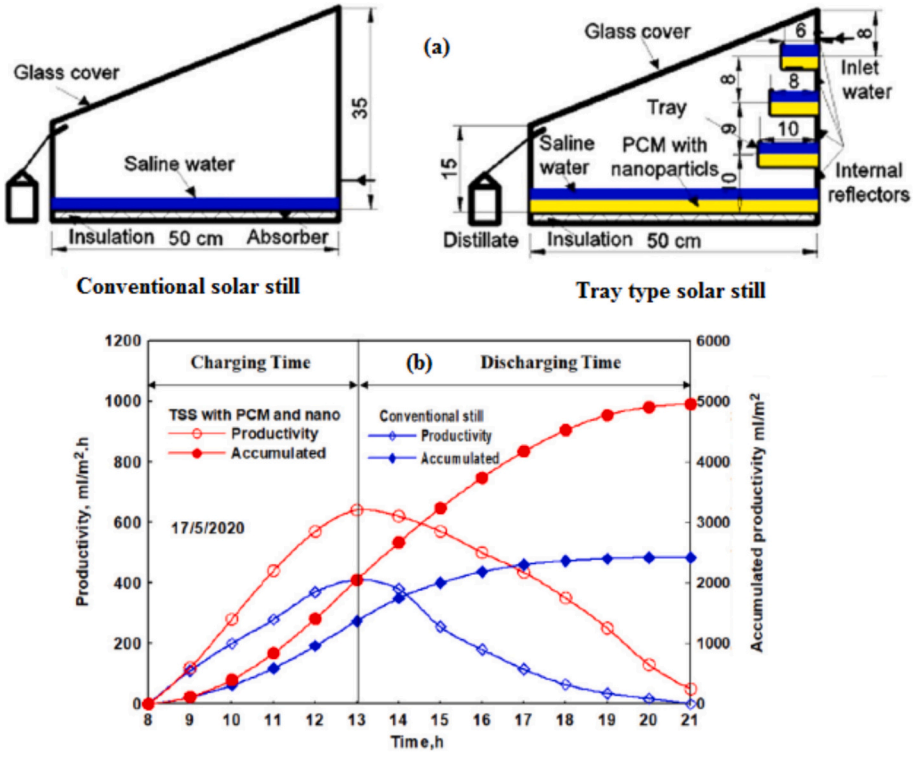


Fig. 31. Schematic of (a) conventional and tray type solar still (b) hourly and accumulated yield using PW [142]. Adapted with permission from Elsevier.

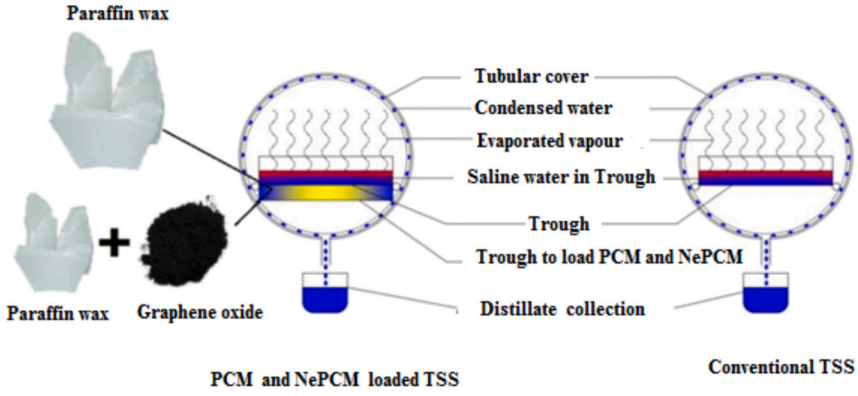


Fig. 32. Schematic of TSS loaded with PCM and NePCM [143].

decreases the heat loss from the side walls and adds radiation to the basin and trays.

The effect of loading PW and NePCM (GO+PW) on tubular solar still (TSS) has been investigated, and rise in a water temperature by ~17.1 % was obtained using NePCM for 0.01 m basin water depth, as represented schematically in Fig. 32 [143]. The NePCM and PCM loaded stills yielded 83.7 % (7.9 kg) and 39.5 % (6.0 kg) higher than the CSS (4.3 kg), with the respective energy efficiencies of 59, 46, and 31 %, while exergy efficiencies were evaluated to be 3.75, 2.2, and 1.67 % from the respective system. Additionally, under the geographical conditions of Chennai (India), TSS loaded with PCM and NePCM has been studied [144]. Upon the addition of GO with a concentration below 0.3 wt%, the conductivity of NePCM was enhanced by 52 % compared to the system loaded with PCM. The daily yields obtained from solar still loaded with NePCM, PCM, and CSS have been explored as 5.62, 3.35, and 2.59 kg, respectively, with the respective energy efficiencies of 50.85, 30.31, and 23.48 % (an increment of 116.5 %).

The effect of adding CuO and Al_2O_3 nano-enhanced PW on single-slope solar still has been investigated [145]. Additionally, CuO was dispersed in the black paint (10 wt%) to enhance the thermal conductivity and coated on the surface of the copper pipe block positioned over the porous material (anthracite) in the basin. The daily yield obtained by various arrangements is shown in Fig. 33. The productivity of solar still has been increased by 55.8 % using CuO-enhanced PW and by 49.5 % using Al_2O_3 enhanced PW at 0.3 wt% concentrations in combination with nano-coated tubes, respectively. Further, a decrease in the melting point of PCM to 2.7 °C and 2.3 °C was reported with the mixing of CuO and Al_2O_3 nanoparticles at 0.3 wt%, respectively. The CuO nanoparticles have been found to be more effective compared to Al_2O_3. Moreover, nano-coating has increased water production costs by about 5.7 %. Selimefendigil et al. [146] have experimented with different single slope solar stills incorporated with paraffin, NePCM (Paraffin+CuO), and NePCM + matt black paint liner, and found productivity enhancement by ~26.8 % using above combined modification. The thermal and exergy efficiency increase was reported by ~4.0 % and 0.75 %, respectively.

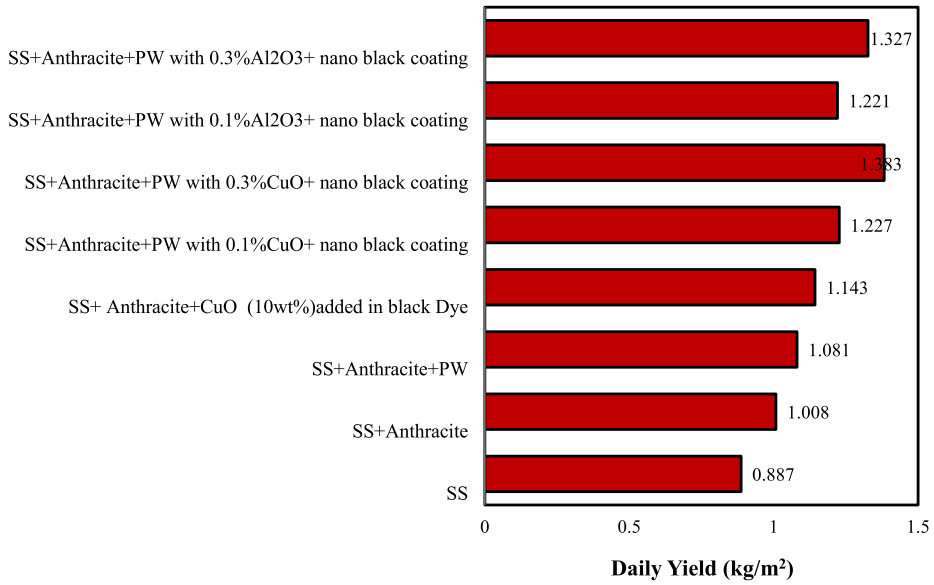


Fig. 33. Variation of yield in double slope solar still incorporated with modifications, derived from [145].

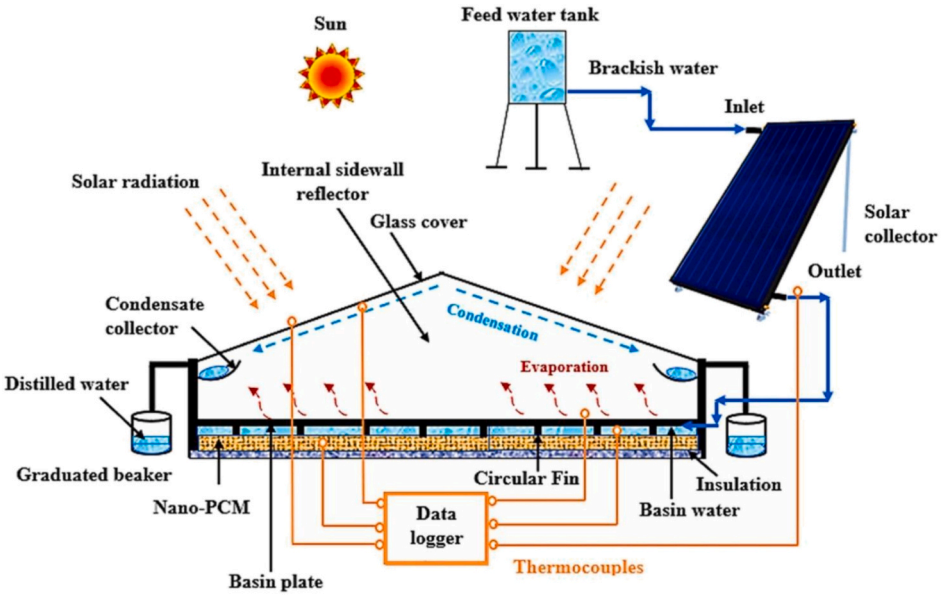


Fig. 34. Modified dual slope still with internal reflector and NePCM [147]. Adapted with permission from Elsevier.

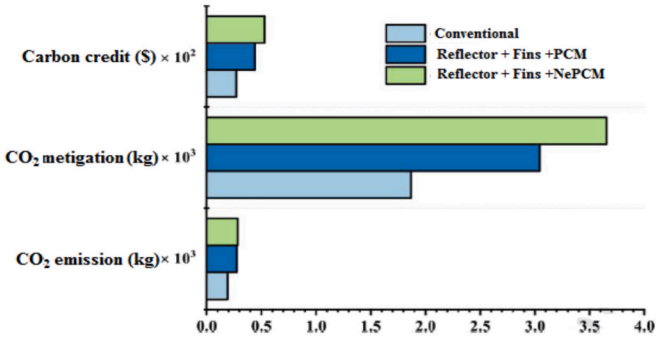


Fig. 35. Comparative environmental effect of adding PCM and NePCM [147].

5.2. *Active solar stills incorporated with NePCM*

Integrating solar stills with collectors, heat pipes, and pulsating heat pipes has been widely used to enhance the yield. However, the performance can be boosted further by incorporating energy storage materials. Dual slope solar still coupled with the collector, internal reflectors, circular fins (hollow), PW, and Al_2O_3 nanoparticles has been experimented as shown in Fig. 34 [147]. The investigation revealed a 51.8 % increase in yield by incorporating the reflector, fins, and PCM combination. The additional increase in yield by 21.5 % has been obtained using NePCM, with an enhancement of thermal and exergy efficiency by 20.1 and 25.0 %, respectively. The distillate production cost decreased by 2 % using NePCM and has been evaluated as 0.020 $/kg with a modified combination, with CO_2 mitigation of 3.65 tons, as depicted in Fig. 35.

Further, the performance of PTC-integrated tubular solar still incorporated with PW-filled aluminum tubes (12 in number) has been investigated using novel PCM-tube geometry, as depicted in Fig. 36

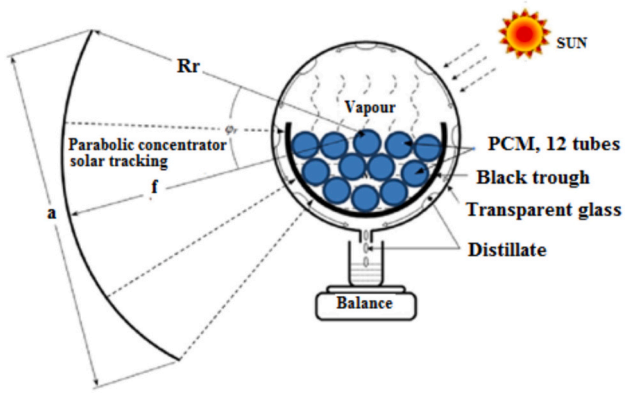


Fig. 36. CPC integrated TSS loaded with NePCM [148]. Adapted with permission from Elsevier.

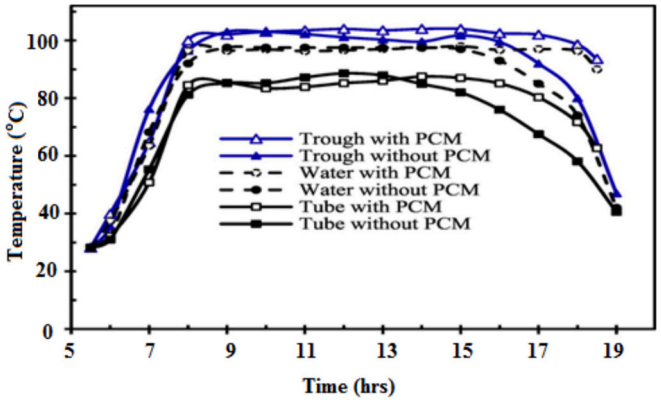


Fig. 37. Temperature profile attainable in various components of solar still during day [148]. Adapted with permission from Elsevier.

[148]. Sun tracking PTC was used to activate the tubular solar still, which consists of 12 aluminum tubes for PCM. The enhancement in productivity by ~40.5 % and energy efficiency by 38.3 % has been estimated. The modified solar still yielded 5.55 kg/m^2 with an energy efficiency of 44.1 % and a production cost of 0.00782 $/kg. The results obtained were higher compared to the yield (3.95 kg/m^2), efficiency (31.9 %), and cost of distillate (0.0163 $/kg) obtained from solar still without PCM. The higher surface temperatures of PCM-filled tubes between 13:00–18:30 h were estimated due to enhanced heat transfer conditions of heterogeneous geometry with large contact surface area with saline water during the low sunshine period (Fig. 37).

Recently, Hammad et al. [32] fabricated a novel design of dome shaped solar still with a stepped conical multi-trays absorbent basin and fountain feeding of saline water. The domed solar distiller has a large surface area compared to a single inclined solar still, and offers enhanced exposure to the ambient leading to reduced condensing cover temperature and improved condensation. The multi-tray stepped conical basin offers a larger area for radiation and increases the vaporization area for water, while fountain feeding minimizes the water thickness in basin trays. The schematic of three techniques used by fabricating the solar still is shown in Fig. 38. The basin was constructed by 0.001 m thick copper sheet of 0.7 m diameter, insulated by 0.004 m thick foam. A mixture of PW with CuO (1 wt%) was stored in one of the still within the conical basin. The findings indicate that daily yield, energy efficiency, and exergy efficiency from NePCM modified solar still were 7.21 kg/m^2, 62.42 %, and 4.16 %, respectively, while the modified solar still without PCM yielded 6.84 kg/m^2 with energy and exergy efficiency of 60.40 % and 3.59 %, respectively. Production cost from the modified solar still with NePCM and without NePCM decreased by 18.48 % and 16.13 %, respectively, compared to the reference solar still without fountain feed (yielded 4.44 kg/m^2). Further, this indicates an augmentation in energy efficiency of 59.61 % and 53.56 %, while exergy efficiency of 123.66 % and 98.39 %, respectively, over the reference solar still.

A summary of studies reported by various investigators related to NePCM-based active solar stills is summarized in Table 1. The use of nanomaterial with PCM has shown a considerable increase in yield and efficiency due to the augmentation of heat transfer coefficient of the composite mixture and a decrease in heat losses from the solar still.

5.3. Miscellaneous NePCM loaded solar stills

5.3.1. Bioactive NePCM loaded solar still

The recent advancements in solar still technology have witnessed a surge in research focusing on nanomaterial integration and novel design configurations. To capture the excess heat and reduce the heat losses to the ambient from the bottom and side walls of SS, a mixture of ricinus communion leaves doped with Ag nanoparticles and paraffin wax was applied in the cavity all around the basin enclosure between the basin and insulation, as shown in Fig. 39 [151]. Solar still was fabricated using a 0.0015 m thick GI sheet of 1 m^2, insulated using plywood and glass wool, and covered from outside by 0.004 m thick transparent glass, allowing for better daytime absorption. The mixture was also employed inside the basin of one of the solar stills. Doping with leaf extract enhances the thermal conductivity and specific heat of the blend. The enhancement of yield/thermal efficiency of three solar stills incorporated with a blend of PW (P) with leaves extract (RC), Ag nanoparticle with leaves extract (ARC), and NePCM with leaves extract (RC) has been found to be 28.5 %/38.3 %, 66 %/44 %, and 96.5 %/48.5 %, respectively, compared to the CSS. The highest yield and thermal efficiency

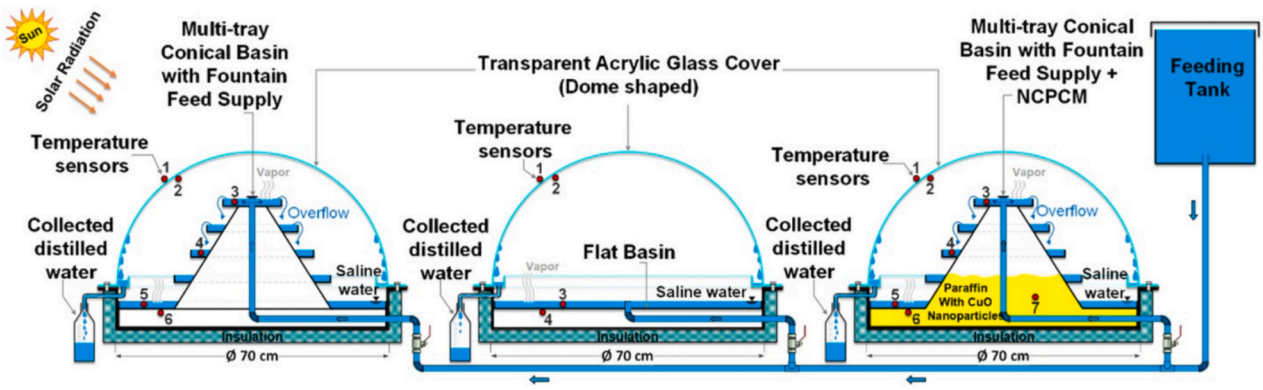


Fig. 38. Schematic of dome shaped solar stills [32]. Adapted with permission from Elsevier.

Table 1
Various studies of 10 years on the effect of NePCM on various configurations of solar stills.

Ref.	Solar still	Country	PCM	Nano-material	Daily yield (kg/m^2)	Outcome
[135]	SS + NePCM	India	PW	Al_2O_3	4.84	• energy efficiency of 66 %.
	SS + PCM		PW	–	4.19	• energy efficiency of 45 %.
[64]	SS + NePCM *Sodium dodeecycl-benzene surfonate	India	SDBS*	Cu	5.28	• yields 35 % higher than PCM alone.
[132]	SS + PCM	Iraq	PW	–	2.64	• yields 10.38 % higher than CCS.
	SS + NePCM		PW	Al_2O_3	3.70	• yields 60.53 % higher than SS+ PCM.
[134]	SS + NePCM	India	PW	TiO_2	4.94	• yields 26 % higher than SS with PW (3.92 kg/m^2)
			PW	CuO	5.28	• yields 34 % higher than SS with PW.
[133]	Stepped + NePCM	Iran	PW	GO	6.25	• yields 25 % higher then still without NePCM.
[137]	SS + NePCM+ reflector	Egypt	PW	Black gravel	3.27	• yields 37.55 % higher than SS with PCM. • Increase in energy and exergy efficiencies by ~37 %.
[138]	SS + NePCM	Egypt	PW	Graphite	8.52	• yields 7.12, 7.47, 7.94, 8.25, and 8.52 kg/m^2 using 0–20 wt% graphite.
[39]	Trays type + NePCM + reflector + nano black paint	Egypt	PW	CuO	12.00	• increased yield and energy efficiency higher by 108 % and 47 %, respectively, than the CSS.
[144]	Tubular +PCM+ GO	Egypt	Polyethylene glycol	GO	5.62	• yields higher by 116.5 % than CSS and 52 % higher than PCM loaded. • energy efficiencies 50.85 %
[140]	SS + NePCM	Iraq	PW	CNTs	5.80	• payback period of 2.3 years • yields 132 % higher than CSS.
	SS + PCM		PW	–	3.50	• payback period of 2.7 years • yields 40 % higher than CSS.
[143]	Tubular + NePCM	India	PW	GO	7.90	• yields 83.7 % higher than CSS.
	tubular +PCM		PW	–	6.00	• yields 39.5 % higher than CSS.
[148]	Tubular + PTC+ aluminum tubes with PCM	Saudi Arabia	PW	*Al* tubes	5.55	• yields 40.5 % higher and energy efficiency by 38.3 % than still without PCM.
	tubular + PTC		–	–	3.95	
[145]	SS + NePCM + nano coating	Iran	PW	Al_2O_3	1.32	• yields 49.5 % higher than CSS.
			PW	CuO	1.38	• yields 55.8 % higher than CSS.
[139]	Dual slope+ NePCM +nanofluid+ Cu chips	Egypt	PW	CuO	6.52	• Efficiency 68 % and CPL 0.011$/kg. • yield, energy, and exergy efficiency enhanced by 113, 112, and 190 %.
[149]	SS + PCM + ferrite material	India	PW	–	5.50	• yields 62 % higher than CSS.
[146]	SS + PCM	Turkey	PW	–	4.10	• nanoparticle increases the yield by 26.77 % than CSS.
	SS + NePCM		PW	CuO	4.53	
	SS + NePCM + black nanopaint		PW	CuO	4.83	
[147]	Dual slope + NePCM +reflector +hollow circular fins +FPC	India	PW	Al_2O_3	3.37	• yield, energy and exergy efficiency higher by 84, 121 and 111 %, respectively. • additional increase in yield by 21.5 % using NePCM than PCM.
[150]	SS + hybrid nano PCM + glass cooling	Iran	PW	Al_2O_3/TiO_2	–	• yield and efficiency increased by 11.09 % and 28.21 %, respectively.
[151]	SS + bioactive PCM	India	PW + RC leaf	–	2.80	• yield and efficiency increased by 28.5 % and 38.3 % than SS.
	SS + bioactive nano		–	Ag + RC leaf	2.98	• yield and efficiency increased 66 % and 44 % higher than SS.
	SS + bioactive NePCM		PW	Ag + RC leaves	3.46	• yield and efficiency increased by 96.5 % and 48.5 % than SS.
[156]	SS + PCM+ hybrid nano material + Magnetic field	Iran	PW	Fe_3O_4 + GO	1.79	• yields 98 % higher than CSS.
	SS + PCM+ hybrid nano material		PW	Fe_3O_4 + GO	1.58	• yields 75 % higher than CSS.
	SS + PCM		PW	–	1.24	• yields 37 % higher than CSS.
[154]	wick SS+ bioactive NePCM	India	PW	TiO_2 + PC leaf	8.70	• yield and efficiency enhanced by 42.74 and 48 %, respectively, than the CSS.
[32]	Domed SS+ stepped conical multi-trays+ fountain	Egypt	PW	CuO	7.21	• Yield improved by 62.38 % than the reference domed still.

increase has been estimated to be enhanced by 124.2 % and 52.0 % using solar still incorporated with all modifications under typical conditions.

Recently, a nano-composite consisting of a mixture of green-sourced *petroselinum crispum* (P) leaf extract blended with TiO_2 nanoparticles (T) and paraffin wax (P) was applied in wick-type solar still [152]. Solar still was made of GI of sloped surface area 1 m^2 and insulated using glass wool. The slope of the surface was 16°, which is equal to the latitude of the place. The inner surface was painted with matt black and contains 3 g of PTP, and a glass cover of 0.004 m thickness was used. The composite samples were applied in the solar still to induce surface plasmon resonance around 450 nm across the absorber plate (Fig. 40). *Petroselinum crispum* leaf offers the benefits of thermal regulation and moisture retention and acts as a reducing and stabilizing agent when mixed with NePCM. The matt-black paint as a nano-composite was applied on the inner surface. The pellets of PTP were placed on the basin surface. The excitation of PTP composition led to significant heat release, attributing to an increased water temperature. The effect of composite concentration shows an optimal yield of 6.721 kg/m^2 during summer and 6.044 kg/m^2 in winter at 30 % concentration, whereas the conventional solar still yielded 2.85 kg/m^2 and 3.160 kg/m^2 in summer and winter, respectively. The night yields of 1.958 kg/m^2 and 1.892 kg/m^2 have been obtained from the modified solar still in summer and winter, respectively. Moreover, conventional solar still yields 0.950 kg/m2 and 0.851 kg/m^2 in summer and winter, respectively. On investigation, it has been revealed that incorporating 30 % nano-composite enhanced

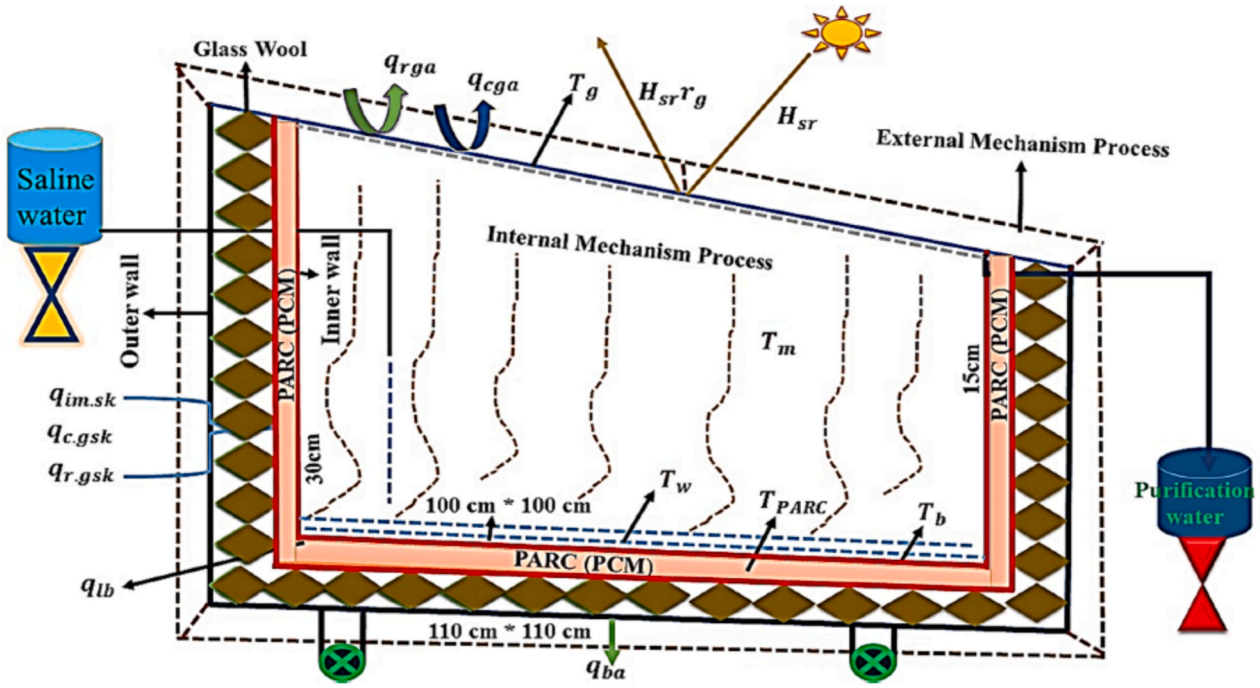


Fig. 39. Schematic of modified solar still [151]. Adapted with permission from Elsevier.

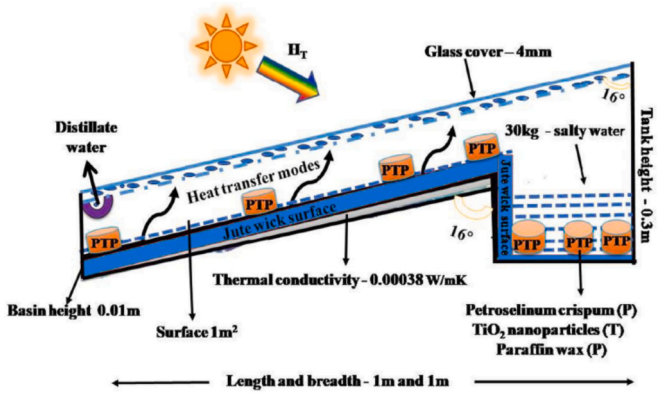


Fig. 40. Schematic of modified wick type solar still [152]. Adapted with permission from Elsevier.

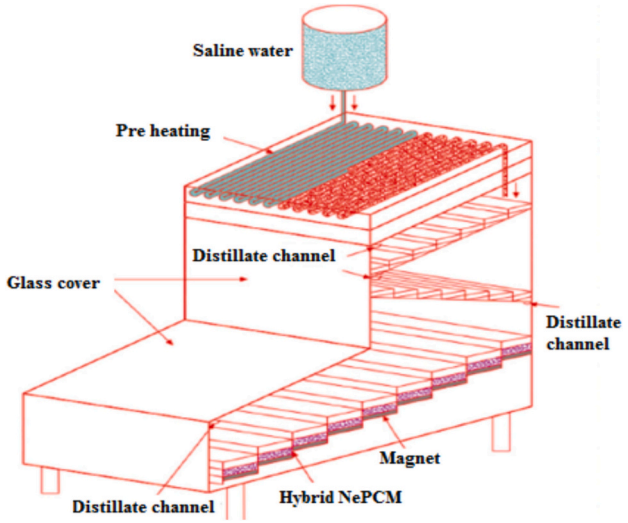


Fig. 41. Modified stepped solar still with hybrid NePCM and magnet [156]. Adapted with permission from Elsevier.

the daily yield extremely (8.70 kg/m^2), including 1.95 kg/m^2 at night on a typical day, resulting in an efficiency of 42.74 %, higher by 48 % than the conventional still.

More recently, Ganeshan [153] used rehashed corn oil as a phase change material to enhance the performance of hybrid PV/T active solar still and found an enhanced PV electrical and thermal efficiency of integrated solar still at 11.7 % and 18.01 %, respectively. The modified solar still produced a daily yield of 7.04 kg/m^2, and corn oil was more effective than coconut, sesame, and peanut oils.

5.3.2. *Hybrid nano materials with PCM and magnetic field*

To study the effect of magnetization in solar still, a mixture of paraffin wax and ferrite material has been incorporated [149]. The total daily yield was found to be enhanced by 62 %, increasing it from 3.4 to 5.5 kg/m^2. The magnetic field breaks the Van der Waals force between the water molecules and soluble salts, increasing the diffusion coefficient and evaporation process. Dhivagar et al. [154] used black iron oxide magnetic powder as an energy storage material to enhance the performance of solar still. The increased daily yield, energy, and exergy efficiency were obtained as 31.2, 18.9, and 19.04 %, respectively. In addition, the still was coupled with the district heating system, and using black powder-coated crushed granite stone significantly improved the distillate yield and was found to be independent of basin water depth [155]. Compared to non-integrated solar still, the enhanced daily yield of 105, 135, and 182 % has been found at 0.01, 0.02, and 0.03 m water depths.

Further, Toosi et al. [156] utilized the effect of the magnetic field in stepped solar still incorporated with a mixture of PW and hybrid nanoparticles (Fe_3O_4 + GO) (Fig. 41). 28 steps were divided into 3 stepped surfaces and made of 0.002 m thick iron sheet. The glass cover thickness was 0.003 m and tilled at 36.26°. Water was preheated using 4 collectors. The magnetic sheets were used under each step of the solar still. Graphene oxide has favorable characteristics and was synthesized and mixed with Fe_3O_4 and PW. The daily yields obtained from solar still

without PCM, solar still with PCM, hybrid NePCM, hybrid NePCM with magnetic field are 0.90 kg, 1.235 kg, 1.577 kg, and 1.79 kg, respectively. The daily yield from solar still incorporated with PCM, hybrid NePCM, and NePCM with the magnetic field have been enhanced by 37, 75, and 98 %, respectively, compared to conventional stepped solar still. The Fe_3O_4 is highly reactive to the magnetic field, which increases the thermal conductivity of NePCM. This is due to the formation of particle chains inside the magnetic nanofluid along the direction of the temperature gradient, lowering the thermal resistance.

5.3.3. Hybrid PCMs loaded solar still

In order to prolong the discharging time of PCMs, a study has been carried out using a hybrid of two PCMs, each having a different melting temperature range [157]. The melting temperature of the PCMs selected has been in the range of 58.03 °C–64.5 °C and 53.05 °C–62 °C, keeping in view that both PCMs should have the same latent heat capacity. However, to increase the discharge time, one PCM in the blend was so selected that it should start discharging once another PCM has almost discharged. An appreciable enhancement yield during daytime/nighttime (2.78/1.62 kg/m^2) has been reported using a solar still incorporated with hybrid PCM compared to the CSS (3.68/00 kg/m^2) and CSS with single PCM (2.72/1.3 kg/m^2). The total accumulated yield (4.40 kg/m^2), energy efficiency (46.29 %), and exergy efficiency (3.52 %) have been found to be higher by ~9 % than the solar still using single PCM.

The use of single organic PCM has many limitations, and to meet the objectives, binary PCMs (eutectic) for energy storage have been developed. The use of eutectic PCMs has shown a significant impact on enhancing the yield of solar stills without increasing the production cost of distillate. Using eutectic PCM (palmitic and stearic acids mixture) and steel wool fiber increased productivity by 35 % while energy efficiency by 37 % [158]. Further, the effect of eutectic PCM (mixture of palmitic/stearic acids, 64/36 wt%) was investigated on dual slope solar still [159]. The eleven copper tubes and six cylinders containing 0.002 kg and 0.009 kg PCM were filled, respectively [Fig. 42]. The energy efficiency was increased by ~37.0 % compared to the conventional design (22.21 %) using binary eutectic PCM.

5.3.4. Hybrid-nanoparticles with PCM

The inclusion of nanofluid boosted the yield of the SS because of the enhanced heat transmission properties of nanoparticles. Recently, a combination of different nanoparticles, such as Al_2O_3/Cu, TiO_2/Ag, TiO_2/Cu, Fe_3O_4/carbon, and CuO/graphene oxide (GO), has been tested to enhance the heat transfer characteristic of thermal systems [123]. The appropriate mixing of hybrid Al_2O_3 + CuO in the ratio of 3:1 with PCM enhanced the thermal conductivity by over 200 % compared to the system operating with pure PCM [56]. Shoeibi et al. [150] explored that combining $Al_2O_3/TiO_2/H_2O$ with an optimal fraction of 0.45 wt% increased the yield and efficiency of solar still by 11.09 % and 28.21 %, respectively. The hybrid Fe_2O_3 and multi-walled carbon nanofluid significantly enhances the base fluid's thermal conductivity and system's energy efficiency [1].

6. Conclusions and future work

6.1. Conclusions

The conventional solar stills are easy to use and inexpensive, but their poor productivity makes them unsuitable for commercial use. An enhanced, consistent productivity can be obtained by incorporating PCMs and NePCMs, reducing production costs and payback time. However, the percentage enhancement depends on various parameters such as the design of solar still, type of PCM, nanoparticle, auxiliary combinations, operational and environmental conditions, etc. The research on various low-cost and eco-friendly PCMs and nanomaterials, including their hybridization, is futuristic. The following are emerging from the review.

a. The theoretical enhancement of yield and energy efficiency for the passive solar still incorporated with stearic acid/basin water ratio of 3:5 have been explored as 80.1 % and 85.3 %, respectively, for typical conditions. Further, incorporating PW in V grooved pyramid-shaped solar still enhanced the yield by 87.4 %. The results revealed that the application of PW is more effective than stearic acid. By incorporating PW, the daytime and nighttime yield enhanced by 7 % and 400 %, respectively, on a typical day.
b. The use of hybrid capric-palmitic acid is better than other PCMs in the order of yield as 100 % > A-48 (92 %) > stearic acid (68 %). However, A-48 PCM is recommended due to its lower cost and environmental benefits.
c. Operating the solar still incorporated with PCM in combination with sensible heat material improves the yield further. CuO-dispersed PW (NePCM) yields 35 % higher than using PW alone. Adding Cu chips in basin water increased the yield, energy, and exergy efficiency by 113, 112, and 190 %, respectively.
d. The productivity enhancement by 30.6, 244, and 415 % has been reported by using CSS+ PW, active design (CSS + PW + FPC), and

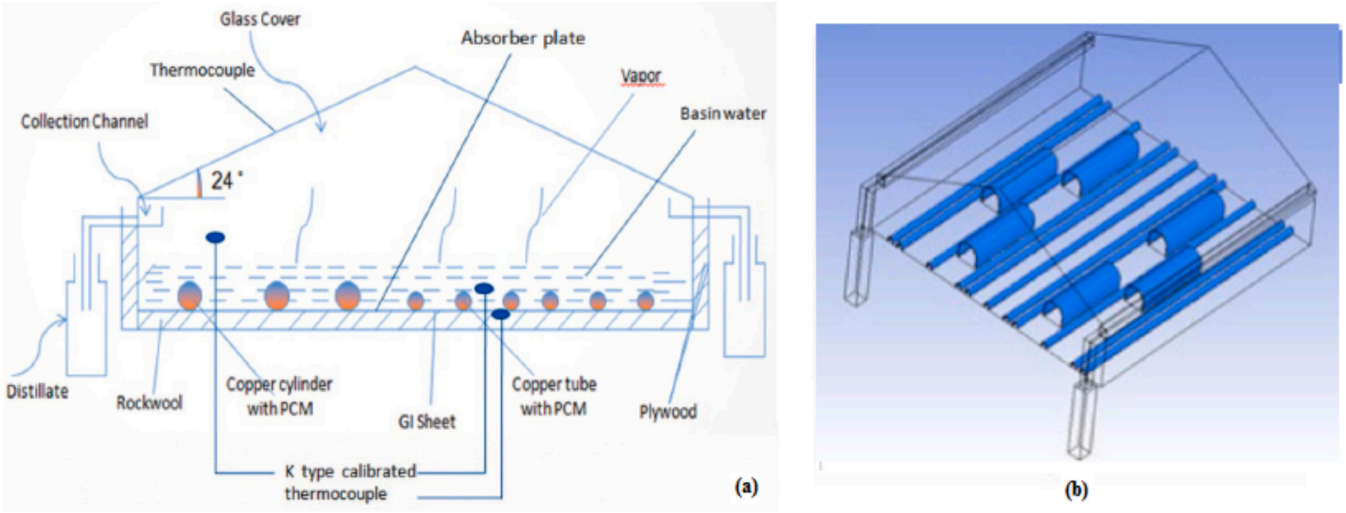


Fig. 42. Schematic of (a) dual slope still with eutectic PCM, (b) PCM tubes and cylinder arrangement [158,159]. Adapted with permission from Elsevier.

CSS+ PW + FPC + UV plasmon resonance solar stills, respectively, compared to the CSS.

e. Dome shaped solar still with stepped conical multi-tray absorbent basin and fountain feed enhanced the performance remarkably, which further enhanced using NePCM.
f. The mixing of nanoparticles has a significant potential for enhancing the thermal conductivity. A mixture of hybrid nanomaterial (Al_2O_3 + CuO) in the ratio of 3:1 enhanced the thermal conductivity of PW by over 200 %, favouring the hybridization with low cost nanomaterial.
g. Rehashed corn oil as PCM in hybrid PV/T active solar still has enhanced electrical and thermal efficiency of integrated solar still by 11.7 % and 18.01 %, respectively, and found to be more effective than coconut, sesame, and peanut oils, with the yield of 7.04 kg/m^2.
h. The daily yield obtained by incorporating internal reflectors+ nano-coating+ NePCM (CuO + PW) enhanced by 5.0 times (~12 kg/m^2) than the CSS (2.4 kg/m^2), with an energy efficiency of 51.5 %.
i. The carbon nanotubes (CNTs) are the most effective nanomaterial, followed by GO, CuO, and Al_2O_3 among the materials under study. The incorporation of NePCM such as CNTs+ PW, GO + PW, CuO+ PW, Al_2O_3+ PW has enhanced the yield by 132 %, 83.7 %, 55.85, and 49.5 %, respectively, than the CSS.
j. Incorporation of bioactive NePCM (Ag + PW+ RC leaves) enhanced the maximum yield and efficiency by 124.2 % and 52.0 %, respectively, compared to the system without PCM.

6.2. Scope for future research

The current review shows that most of the studies have been carried out using paraffin wax as PCM. Moreover, limited studies are available on bioactive NePCM, bio-based PCM, hybrid PCMs, and hybrid NePCMs. The further recommendations are;

(i) Wicked prismatic-shaped basin solar still and dome-shaped spherical solar still (stepped conical multi-trays absorbent basin and fountain feed) have the potential. Further investigations should be carried out incorporating various low cost NePCMs and feeding of preheated high temperature saline water using collectors/concentrators to compete with the existing configurations of solar stills.
(ii) The optimization of NePCM loaded solar stills in combination with other sensible heat storage materials can be the subject of future studies from the viewpoint of thermal management.
(iii) Synthesizing of various nanomaterials with different PCMs to investigate the impact of their utilization on the performance.
(iv) Future investigations are suggested to study the performance using plasmon resonance and magnetic field, which has the potential in combination with NePCMs.
(v) Explore the potential of advanced machine learning algorithms for forecasting the performance of various configurations of solar stills under various geographical locations.

CRediT authorship contribution statement

Aseem Dubey: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Akhilesh Arora:** Writing – review & editing, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

The authors express heartfelt thanks to Delhi Technological University (DTU), Delhi to provide required facilities for carrying out this work.

References

[1] S.D. Kanka, M.G. Kibria, U.A. Anika, B.K. Das, M.S. Hossain, D. Roy, M. S. Mohtasim, Impact of various environmental parameters and production enhancement techniques on direct solar still: a review, Sol. Energy 267 (2024) 112216.
[2] T. Luo, R. Young, P. Reig, Aqueduct projected water stress country rankings. World Resources Institute (Technical Note), WRI. https://www.wri.org/data/aqueduct-projected-water-stress-country-rankings, 2015.
[3] The Water Gap State of Water report lr pages.pdf. Water Aid, The water gap, 2018, pp. 1–24. https://washmatters.wateraid.org/sites/g/files/jkxoof256/files/.
[4] L. Zhang, Z. Xu, L. Zhao, B. Bhatia, Y. Zhong, S. Gong, E.N. Wang, Passive, high-efficiency thermally-localized solar desalination, Energy Environ. Sci. 14 (2021) 1771–1793.
[5] S.M. Parsa, A. Rahbar, Y.D. Javadi, M.H. Koleini, M. Afrand, M. Amidpour, Energy-matrices, exergy, economic, environmental, exergoeconomic, nviroeconomic, and heat transfer (6E/HT) analysis of two passive/active solar still water desalination nearly 4000m: altitude concept, J. Clean. Prod. 261 (2020) 121243.
[6] A. Elsheikh, M. Zayed, A. Aboghazala, F.A. Essa, S. Rehman, O.L. Muskens, A. Kamal, M.A. Elaziz, Innovative solar distillation system with prismatic absorber basin: experimental analysis and LSTM machine learning modeling coupled with great wall construction algorithm, Process. Saf. Environ. Prot. 186 (2024) 1120–1133.
[7] V.K. Sonker, J.P. Chakraborty, A. Sarkar, Development of a frugal solar still using phase change material and nanoparticles integrated with commercialization through a novel economic model, J. Energy Storage 51 (2022) 104569.
[8] S.K. Hota, S.S. Hada, C. Keske, G. Diaz, Feasibility of desalination by solar stills for small community scale freshwater demand, J. Clean. Prod. 379 (1) (2022) 134595.
[9] A.E. Kabeel, K. Harby, M. Abdelgaied, A. Eisa, A comprehensive review of tubular solar still designs, performance, and economic analysis, J. Clean. Prod. 246 (2019) 119030.
[10] M.M. Younes, A.S. Abdullah, Z.M. Omara, F.A. Essa, Enhancement of discs' solar still performance using thermal energy storage unit and reflectors: an experimental approach, Alex. Eng. J. 61 (10) (2022) 7477–7487.
[11] G. Xiao, X. Wang, M. Ni, F. Wang, W. Zhu, Z. Luo, K. Cen, A review on solar stills for brine desalination, Appl. Energy 103 (2013) 642–652.
[12] H.E.S. Fath, M. El-Samanoudy, K. Fahmy, A. Hassabou, Thermal-economic analysis and comparison between pyramid-shaped and single-slope solar still configurations, Desalination 159 (2003) 69–79.
[13] A.J.N. Khalifa, On the effect of cover tilt angle of the simple solar still on its produc- tivity in different seasons and latitudes, Energy Convers. Manag. 52 (1) (2011) 431–436.
[14] S.S. Tuly, M.S. Rahman, M.R.I. Sarker, R.A. Beg, Combined influence of fin, phase change material, wick, and external condenser on the thermal performance of a double slope solar still, J. Clean. Prod. 287 (2021) 125458.
[15] S.S.A. Toosi, H.R. Goshayeshi, S.Z. Heris, Experimental investigation of stepped solar still with phase change material and external condenser, J. Energy Storage 40 (2021) 102681.
[16] H. Amiri, M. Aminy, M. Lotfi, B. Jafarbeglo, Energy and exergy analysis of a new solar still composed of parabolic trough collector with built-in solar still, Renew. Energy 163 (2021) 465–479.
[17] Q. Huang, C. Du, C. Guo, C. Huang, X. Wang, A high-efficiency salt-rejecting solar evaporator with optimized porous structure for continuous solar desalination, Appl. Therm. Eng. 187 (2021) 116515.
[18] B. Omidi, N. Rahbar, H. Kargarsharifabad, S. Rashidi, Combination of a solar collector and thermoelectric cooling modules in a humidification–dehumidification desalination system-experimental investigation with energy, exergy, exergoeconomic and environmental analysis, Energy Convers. Manag. 225 (2020) 113440.
[19] P. Azari, A.M. Lavasani, N. Rahbar, Evaluation of energy and exergy efficiency of a solar air collector with V grove absorber plate using mathematical modeling, Iran J. Mech. Eng. 22 (2) (2020) 148–175.
[20] P. Azari, A.M. Lavasani, N. Rahbar, M.E. Yazdi, Performance enhancement of a solar still using a V-groove solar air collector—experimental study with energy, exergy, enviroeconomic, and exergoeconomic analysis, Environ. Sci. Pollut. Res. 28 (2021) 65525–65548.
[21] P. Azari, A.M. Lavasani, N. Rahbar, M.E. Yazdi, Combination of a v-grooved solar collector with a single slope solar still: performance evaluation, mathematical

modeling, and economic analysis, Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. 236 (15) (2022) 8731–8753.
[22] S. Kumar, G.N. Tiwari, Estimation of internal heat transfer coefficients of a hybrid (PV/T) active solar still, Sol. Energy 83 (2009) 1656–1667.
[23] S. Shoeibi, M. Saemian, M. Khiadani, H. Kargarsharifabad, S. Ali Agha Mirjalily, Influence of PV/T waste heat on water productivity and electricity generation of solar stills using heat pipes and thermoelectric generator: an experimental study and environmental analysis, Energy Convers. Manag. 276 (2023) 116504.
[24] S.K. Suraparaju, S.K. Natarajan, Waste Tyre pieces as photothermal absorbers for efficient evaporation in solar desalination systems, Clean Techn. Environ. Policy 25 (2023) 3161–3176.
[25] N. Rahbar, J.A. Esfahani, A. Asadi, An experimental investigation on productivity and performance of a new improved design potable asymmetrical solar still utilizing thermoelectric modules, Energy Convers. Manag. 118 (2016) 55–62.
[26] Z.M. Omara, A.S. Abdullah, A.E. Kabeel, F.A. Essa, The cooling techniques of the solar stills' glass covers-a review, Renew. Sust. Energ. Rev. 78 (2017) 176–193.
[27] S. Shoeibi, N. Rahbar, A.A. Esfahlani, H. Kargarsharifabad, Energy matrices, economic and environmental analysis of thermoelectric solar desalination using cooling fan, J. Therm. Anal. Calorim. 147 (2022) 9645–9660.
[28] S.K. Suraparaju, S.K. Natarajan, Effect of natural sisal fibre on enhancing the condensation rate of solar still for sustainable clean water production, Therm. Sci. Eng. Prog. 36 (2022) 101527.
[29] S. Nazari, H. Safarzadeh, M. Bahiraei, Experimental and analytical investiga tions of productivity, energy and exergy efficiency of a single slope solar still enhanced with thermoelectric channel and nanofluid, Renew. Energy 135 (2019) 729–744.
[30] S. Shoeibi, H. Kargarsharifabad, N. Rahbar, G. Khosravi, M. Sharifpur, An integrated solar desalination with evacuated tube heat pipe solar collector and new wind ventilator external condenser, Sustain Energy Technol Assess 50 (2022) 101857.
[31] F.A. Hammad, S.M. Shalaby, M.E. Zayed, Innovative design of a dome-shaped solar distiller enhanced with a multi-tray conical absorber basin and fountain feeding supply: experimental study with energy, exergy, and enviro-economic analyses, Sol. Energy 266 (2023) 112183.
[32] F.A. Hammad, M.E. Zayed, E. El-Bialy, K. Irshad, Z. Lu, S.M. Shalaby, Performance assessment of a novel solar distiller with a fountain-shaped basin design embedded with phase change materials enriched with copper oxide nano-additives: a detailed experimental investigation, J. Energy Storage 82 (2024) 110555.
[33] H.K.S. Abad, M. Ghiasi, S.J. Mamouri, M.B. Shafii, A novel integrated solar desalination system with a pulsating heat pipe, Desalination 311 (2013) 206–210.
[34] N. Rahbar, J.A. Esfahani, Experimental study of a novel potable solar still by utilizing the heatpipe and thermoelectric module, Desalination 284 (2012) 55–61.
[35] T. Rajaseenivasan, K.K. Murugavel, T. Elango, Performance and exergy analysis of a double-basin solar still with different materials in basin, Desalin. Water Treat. 55 (2015) 1786–1794.
[36] A.E. Kabeel, S.A. El-Agouz, R. Sathyamurthy, T. Arunkumar, Augmenting the productivity of solar still using jute cloth knitted with sand heat energy storage, Desalination 443 (2018) 122–129.
[37] R. Dhivagar, S. Shoeibi, S.M. Parsa, S. Hoseinzadeh, H. Kargarsharifabad, M. Khiadani, Performance evaluation of solar still using energy storage biomaterial with porous surface: an experimental study and environmental analysis, Renew. Energy 206 (2023) 879–889.
[38] S.K. Suraparaju, R. Dhanusuraman, S.K. Natarajan, Performance evaluation of single slope solar still with novel pond fibres, Process. Saf. Environ. Prot. 154 (2021) 142–154.
[39] A.S. Abdullah, Z.M. Omara, F.A. Essa, U.F. Alqsair, M. Aljaghtham, I.B. Mansir, S. Shanmugan, W.H. Alawee, Enhancing trays solar still performance using wick finned absorber, nano- enhanced PCM, Alex. Eng. J. 61 (12) (2022) 12417–12430.
[40] S. Shoeibi, M. Saemian, H. Kargarsharifabad, S. Hosseinzade, N. Rahbar, M. Khiadani, M.M. Rashidi, A review on evaporation improvement of solar still desalination using porous material, Int. Commun. Heat Mass Transf. 138 (2022) 106387.
[41] M.E.H. Attia, A.E. Kabeel, M.E. Zayed, M. Arıcı, M. Abdelgaied, Optimal size of spherical rock salt balls as low-cost thermal storage materials for performance augmentation of hemispherical solar distillers: experimental investigation and thermo-economic analysis, J. Clean. Prod. 374 (2022) 134006.
[42] A.E. Kabeel, R. Sathyamurthy, S.W. Sharshir, A. Muthumanokar, H. Panchal, N. Prakash, C. Prasad, S. Nandakumar, M.S.E. Kady, Effect of water depth on a novel absorber plate of pyramid solar still coated with TiO_2 nano black paint, J. Clean. Prod. 213 (2019) 185–191.
[43] S. Nazari, H. Safarzadeh, M. Bahiraei, Performance improvement of a single slope solar still by employing thermoelectric cooling channel and copper oxide nanofluid: an experimental study, J. Clean. Prod. 208 (2019) 1041–1052.
[44] W. Chen, C. Zou, X. Li, H. Liang, Application of recoverable carbon nanotube nanofluids in solar desalination system: an experimental investigation, Desalination 451 (2019) 92–101.
[45] L.D. Jathar, S. Ganesan, Assessing the performance of concave type stepped solar still with nanoparticles and condensing cover cooling arrangement: an experimental approach, Groundw. Sustain. Dev. 12 (2021) 100539.
[46] S. Shoeibi, H. Kargarsharifabad, S.A.A. Mirjalily, M. Sadi, A. Arabkoohsar, A comprehensive review of nano-enhanced phase change materials on solar energy applications, J. Energy Storage 50 (2022) 104262.
[47] M. Abdelgaied, M.E.H. Attia, A.E. Kabeel, M.E. Zayed, Improving the thermo-economic performance of hemispherical solar distiller using copper oxide nanofluids and phase change materials: experimental and theoretical investigation, Sol. Energy Mater. Sol. Cells 238 (2022) 111596.
[48] M.R. Rajamanickam, A. Ragupathy, Influence of water depth on internal heat and mass transfer in a double slope solar still, Energy Procedia 14 (2012) 1701–1708.
[49] M.A. Arabi, Y. Zurigat, H.A. Hinai, S.A. Hiddabi, Modeling and performance analysis of a solar desalination unit with double-glass cover cooling, Desalination 143 (2) (2002) 173–182.
[50] K. Selvaraj, A. Natarajan, Factors influencing the performance and productivity of solar stills - a review, Desalination 435 (2018) 181–187.
[51] S. Shoeibi, S.A.A. Mirjalily, H. Kargarsharifabad, H. Panchal, R. Dhivagar, Comparative study of double-slope solar still, hemispherical solar still, and tubular solar still using Al_2O_3/water film cooling: a numerical study and CO_2 mitigation analysis, Environ. Sci. Pollut. Res. 29 (2022) 65353–65369.
[52] S. Shoeibi, S.A.A. Mirjalily, H. Kargarsharifabad, M. Khiadani, H. Panchal, A comprehensive review on performance improvement of solar desalination with applications of heat pipes, Desalination 540 (2022) 115983.
[53] P.H. Niknam, A. Sciacovelli, Hybrid PCM-steam thermal energy storage for industrial processes – link between thermal phenomena and techno-economic performance through dynamic modeling, Appl. Energy 331 (2023) 120358.
[54] H. Mousa, A.M. Gujarathi, Modeling and analysis the productivity of solar desalination units with phase change materials, Renew. Energy 95 (2016) 225–232.
[55] A. Radhwan, Transient performance of a steeped solar still with built-in latent heat thermal energy storage, Desalination 171 (1) (2005) 61–76.
[56] S.M. Shalaby, E.E. Bialy, A.A.E. Sebaii, An experimental investigation of a v-corrugated absorber single-basin solar still using PCM, Desalination 398 (2016) 247–255.
[57] A.E. Kabeel, M. Abdelgaied, Improving the performance of solar still by using PCM as a thermal storage medium under Egyptian conditions, Desalination 383 (2016) 22–28.
[58] A.A.E. Sebaii, A.A.A. Ghamdi, F.S.A. Hazmi, A.S. Faidah, Thermal performance of a single basin solar still with PCM as a storage medium, Appl. Energy 86 (7–8) (2009) 1187–1195.
[59] A.E. Kabeel, M. Abdelgaied, M. Mahgoub, The performance of a modified solar still using hot air injection and PCM, Desalination 379 (2016) 102–107.
[60] P. Khalilmoghadam, A. Rajabi-Ghahnavieh, M.B. Shafii, A novel energy storage system for latent heat recovery in solar still using phase change material and pulsating heat pipe, Renew. Energy 163 (2021) 2115–2127.
[61] A.E. Kabeel, Y.A.F. El-Samadony, W.M. El-Maghlany, Comparative study on the solar still performance utilizing different PCM, Desalination 432 (2018) 89–96.
[62] H. Hafs, M. Asbik, H. Boushaba, A. Koukouch, A. Zaaoumi, A. Bah, O. Ansari, Numerical simulation of the performance of passive and active solar still with corrugated absorber surface as heat storage medium for sustainable solar desalination technology, Groundw. Sustain. Dev. 14 (2021) 100610.
[63] S.L. Tariq, H.F. Alib, M.A. Akramc, M.M. Janjuad, M. Ahmadlouydarabe, Nanoparticles enhanced phase change materials (NePCMs)-a recent review, Appl. Therm. Eng. 176 (2020) 115305.
[64] M.T. Chaichan, H.A. Kazem, Using aluminium powder with PCM (paraffin wax) to enhance single slope solar water distillation productivity in Baghdad-Iraq winter weathers, Int. J. Renew. Energy Res. 5 (1) (2015) 251–257.
[65] D.D.W. Rufuss, S. Iniyan, L. Suganthi, P.A. Davies, Low mass fraction impregnation with graphene oxide (GO) enhances thermo-physical properties of paraffin for heat storage applications, Thermochim. Acta 655 (2017) 226–233.
[66] J. Karthikeyan, P. Selvaraj, G. Nagaraj, Day and night yield performance analysis of solar still for saline water using energetic materials with thermocol insulation, Mater. Today: Proc. 33 (2020) 4848–4851.
[67] C.J. Ho, T.Y. Gao, Preparation and thermophysical properties of nanoparticle in paraffin emulsion as phase change material, Int. Commun. Heat Mass Transf. 36 (2009) 467–470.
[68] J. Wang, H. Xie, Z. Guo, Y. Li, Improved thermal properties of paraffin wax by the addition of TiO_2 nanoparticles, Appl. Therm. Eng. 73 (2014) 1541–1547.
[69] Y. Tang, D. Su, X. Huang, G. Alva, L. Liu, G. Fang, Synthesis and thermal properties of the MA/HDPE composites with nano-additives as form-stable PCM with improved thermal conductivity, Appl. Energy 180 (2016) 116–129.
[70] A. Elfasakhany, Performance assessment and productivity of a simple-type solar still integrated with nanocomposite energy storage system, Appl. Energy 183 (2016) 399–407.
[71] A. Sciacovelli, F. Colella, V. Verda, Melting of PCM in a thermal energy storage unit: numerical investigation and effect of nanoparticle enhancement, Int. J. Energy Res. 37 (2013) 1610–1623.
[72] K. Biswas, J. Lu, P. Soroushian, S. Shrestha, Combined experimental and numerical evaluation of a prototype nano-PCM enhanced wallboard, Appl. Energy 131 (2014) 517–529.
[73] F.A. Essa, Z. Omara, A. Abdullah, S. Shanmugan, H. Panchal, A.E. Kabeel, R. Sathyamurthy, M.M. Athikesavan, A. Elsheikh, M. Abdelgaied, B. Saleh, Augmenting the productivity of stepped distiller by corrugated and curved liners, CuO/paraffin wax, wick, and vapor suctioning, Environ. Sci. Pollut. Res. 28 (2021) 56955–56965.
[74] A.A. Saeed, A.M. Alharthi, K.M. Aldosari, A.S. Abdullah, F.A. Essa, U.F. Alqsair, M. Aljaghtham, Z.M. Omara, Improving the drum solar still performance using corrugated drum and nano-based phase change material, J. Energy Storage 55 (C) (2022) 105647.
[75] D.D.W. Rufuss, S. Arulvel, S. Iniyan, L. Suganthi, Numerical study of titanium oxide nanoparticle enhanced energy storage material in solar desalination, Mater. Today Proc. 43 (2) (2021) 805–808.

[76] M.M.A. Saeed, D.M. Hachim, H.G. Hameed, Numerical investigation for single slope solar still performance with optimal amount of nano-PCM, 12/23, J. Adv. Res. Fluid Mech. Therm. Sci. 63 (2) (2020) 302–316.
[77] S. Moreno, C. Alvarez, J.F. Hinojosa, V.M. Maytorena, Numerical analysis of a solar still with phase change material under the basin, J. Energy Storage 55 (2022) 105427.
[78] S. Shoeibi, H. Kargarsharifabad, N. Rahbar, G. Ahmadi, M.R. Safaei, Performance evaluation of a solar still using hybrid nanofluid glass cooling-CFD simulation and environmental analysis, Sustain Energy Technol Assess 49 (2022) 101728.
[79] F. Jamil, F. Hassan, S. Shoeibi, M. Khiadani, Application of advanced energy storage materials in direct solar desalination: a state of art review, Renew. Sust. Energ. Rev. 186 (2023) 113663.
[80] A. Hemmatian, H. Kargarsharifabad, A.A. Esfahlani, N. Rahbar, S. Shoeibi, Improving solar still performance with heat pipe/pulsating heat pipe evacuated tube solar collectors and PCM: an experimental and environmental analysis, Sol. Energy 269 (2024) 112371.
[81] P. Devaux, M.M. Farid, Benefits of PCM under floor heating with PCM wallboards for space heating in winter, Appl. Energy 191 (2017) 593–602.
[82] W. Wang, X. Yang, Y. Fang, J. Ding, Preparation and performance of form-stable polyethylene glycol/silicon dioxide with heat storage system using phase change material (PCM), Desalination 381 (2009) 26–37.
[83] L. Wang, D. Meng, Fatty acid eutectic/polymethyl methacrylate composite as form-stable phase change material for thermal energy storage, Appl. Energy 87 (8) (2010) 2660–2665.
[84] O. Okogeri, V.N. Stathopoulos, What about greener phase change materials? A review on biobased phase change materials for thermal energy storage applications, Thermofluids 10 (2021) 100081.
[85] W. Su, J. Darkwa, G. Kokogiannakis, Review of solid–liquid phase change materials and their encapsulation technologies, Renew. Sust. Energ. Rev. 48 (2015) 373–391.
[86] R. Gugulothu, N.S. Somanchi, G. Devendar, P.K.N. Deepika Vilasagarapu, Solar water distillation using different phase change materials, Mater. Today: Proc. 4 (2) (2017) 314–321.
[87] M. Shashikanth, B. Khadka, Y. Lekhana, P.M.S. Kiran, N. Alaparthi, S. Veerammneni, Solar water distillation using energy storage material, Prog Earth Planet Sci 11 (2015) 368–375.
[88] P.K.S. Rathore, S.K. Shukla, N.K. Gupta, Yearly analysis of peak temperature, thermal amplitude, time lag and decrement factor of a building envelope in tropical climate, J. Build. Eng. 31 (2020) 101459.
[89] B.P. Jelle, S.E. Kalnæs, Phase change materials for application in energy-efficient buildings, in: F.P. Torgal, C.G. Granqvist, B.P. Jelle, G.P. Vanoli, N. Bianco, J. Kurnitski (Eds.), Cost-Effective Energy Efficient Building Retrofitting, Elsevier Ltd, 2017, pp. 57–118.
[90] Y. Zhang, X. Li, J. Li, C. Ma, L. Guo, X. Meng, Solar driven phase change micro-encapsulation with efficient Ti_4O_7 nanoconverter for latent heat storage, Nano Energy 53 (2018) 579–586.
[91] A. Vasu, F.Y. Hagos, M.M. Noor, R. Mamat, W.H. Azmi, A.A. Abdullah, T.K. I. Ansari, Corrosion effect of phase change materials in solar thermal energy storage application, Renew. Sust. Energ. Rev. 76 (2017) 19–33.
[92] S. Kumar, A. Dubey, G.N. Tiwari, A solar still augmented with an evacuated tube collector in forced mode, Desalination 347 (2014) 15–24.
[93] G.N. Tiwari, A. Tiwari, Solar Distillation Practice for Water Desalination Systems, first ed., Anamaya Publications, New Delhi, 2007.
[94] F. Sarhaddi, F.F. Tabrizi, H.A. Zoori, S.A.H.S. Mousavi, Comparative study of two weir type cascade solar stills with and without PCM storage using energy and exergy analysis, Energy Convers. Manag. 133 (2017) 97–109.
[95] M. Asbik, O. Ansari, A. Bah, N. Zari, A. Mimet, H.E. Ghetany, Exergy analysis of solar desalination still combined with heat storage system using phase change material (PCM), Desalination 381 (2016) 26–37.
[96] M. Dashtban, F. Tabrizi, Thermal analysis of a weir-type cascade solar still integrated with PCM storage, Desalination 279 (1–3) (2011) 415–422.
[97] M. Kumar, C. Yadav, H. Manchanda, Combined effects of composite thermal energy storage and magnetic field to enhance productivity in solar desalination, I. Jr. Adv. Res. Innov. 4 (1) (2016) 335–342.
[98] A.E. Kabeel, M.A. Teamah, M. Abdelgaied, G.B.A. Aziz, Modified pyramid solar still with v-corrugated absorber plate and PCM as a thermal storage medium, J. Clean. Prod. 161 (2017) 881–887.
[99] V.S. Vigneswarn, P.G. Kumar, D. Sakthivadivel, K. Balaji, M. Meikandan, B. V. Dinakar, K.K. Kamal, G. Kumaresan, Energy, exergy, and economic analysis of low thermal conductivity basin solar still integrated with phase change material for energy storage, J. Energy Storage 34 (2021) 102194.
[100] M.S. Yousef, H. Hassan, Energetic and exergetic performance assessment of the inclusion of phase change materials (PCM) in a solar distillation system, Energy Convers. Manag. 179 (2019) 349–361.
[101] J. Kateshia, V.J. Lakhera, Analysis of solar still integrated with phase change material and pin fins as absorbing material, J. Energy Storage 35 (2021) 102292.
[102] A.K. Thakur, P. Khandelwal, B. Sharma, Productivity comparison of solar still with nano fluid and phase changing material with same depth of water, in: G. Anand, J.K. Pandey, S. Rana (Eds.), Nanotechnology for Energy and Water, Springer, Switzerland, 2017, pp. 119–129.
[103] S.W. Sharshir, G. Peng, L. Wu, N. Yang, F.A. Essa, A.H. Elsheikhd, I.T.M. Showgi, A.E. Kabeel, Enhancing the solar still performance using nanofluids and glass cover cooling: experimental study, Appl. Therm. Eng. 113 (2017) 684–693.
[104] S.W. Sharshir, G. Peng, L. Wu, F.A. Essa, A.E. Kabeel, N. Yang, The effects of flake graphite nanoparticles, phase change material, and film cooling on the solar still performance, Appl. Energy 191 (2017) 358–366.
[105] W.L. Cheng, Y.K. Huo, Y.L. Nian, Performance of solar still using shape-stabilized PCM: experimental and theoretical investigation, Desalination 455 (2019) 89–99.
[106] R. Sathyamurthy, P. Nagarajan, J. Subramani, D. Vijayakumar, K.M.A. Ali, Effect of water mass on triangular pyramid solar still using phase change material as storage medium, Energy Procedia 61 (2014) 2224–2228.
[107] S. Ravishankar, P.K. Nagarajan, D. Vijayakumar, M.K. Jawahar, Phase change material on augmentation of fresh water production using pyramid solar still, Int. J. Renew. Energy Dev. 2 (3) (2013) 115–120.
[108] A.K. Rai, V. Sachan, Experimental study of a tubular solar still with phase change material, Int. J. Mech. Eng. Technol. 6 (1) (2015) 42–46.
[109] K. Swetha, J. Venugopal, Experimental investigation of a single slope solar still using PCM, Int. J. Res. Environ. Sci Technol. 1 (2011) 30–33.
[110] A.A.F. Al-Hamadani, S.K. Shukla, Modelling of solar distillation system with phase change materials to rage medium, Therm. Sci. 18 (2) (2014) 347–362.
[111] T. Arunkumar, D. Denkenberger, A. Ahsan, R. Jayaprakash, The augmentation of distillate yield by using concentrator coupled solar still with phase change material, Desalination 314 (2013) 189–192.
[112] A. Kabeel, M. Abdelgaied, Observational study of modified solar still coupled with oil serpentine loop from cylindrical parabolic concentrator and phase changing material under basin, Sol. Energy 144 (2017) 71–78.
[113] M. Faegh, M.B. Shafii, Experimental investigation of a solar still equipped with an external heat storage system using phase change materials and heat pipes, Desalination, 409, 128–135.
[114] M. Al-Harahsheh, M. Abu-Arabi, H. Mousa, Z. Alzghoul, Solar desalination using solar still enhanced by external solar collector and PCM, Appl. Therm. Eng. 128 (2018) 1030–1040.
[115] E.H. Mehdiabadi, F. Sarhaddi, F. Sobhnamayan, Exergy performance evaluation of a basin-type double-slope solar still equipped with phase-change material and PV/T collector, Renew. Energy 145 (2020) 2409–2425.
[116] V.K. Sonker, J.P. Chakraborty, A. Sarkar, R.K. Singh, Solar distillation using three different phase change materials stored in a copper cylinder, Energy Rep. 5 (2019) 1532–1542.
[117] A.E. Mazraeh, M. Babayan, M. Yari, A.M. Sefidan, S.C. Saha, Theoretical study on the performance of a solar still system integrated with PCM-PV module for sustain- able water and power generation, Desalination 443 (2018) 184–197.
[118] A.H. Abed, H.A. Hoshi, M.H. Jabal, Experimental investigation of modified solar still coupled with high-frequency ultrasonic vaporizer and phase change material capsules, Case Stud. Therm. Eng. 28 (2021) 101531.
[119] T. Arunkumar, A.E. Kabeel, Effect of phase change material on concentric circular tubular solar still-integration meets enhancement, Desalination 414 (2017) 46–50.
[120] M. Gowtham, K.R. Neiel, V. Nagarajan, P.C. Dass, A. Thimothy, Integrated performance analysis of latent heat storage and finned type solar distiller, Int. J. Eng. Technol. 4 (5) (2012) 613–616.
[121] M.S. Chander, M. Gowtham, Concentrated parabolic solar distiller integrated with latent heat storage material and mini solar pond, Int. J. Chem. Eng. App. 2 (3) (2011) 2–7.
[122] A. Amarloo, M. Shafii, Enhanced solar still condensation by using a radiative cooling system and phase change material, Desalination 467 (2019) 43–50.
[123] H. Ajdari, A. Ameri, Performance assessment of an inclined stepped solar still integrated with PCM and CuO/GO nanocomposite as a nanofluid, J. Build. Eng. 49 (2022) 104090.
[124] P.R. Prasad, P. Pujitha, B.G.V. Rajeev, K. Vikky, Energy efficient solar water still, Chem. Tech. Res. 3 (4) (2011) 1781–1787.
[125] J.M. Khodadadi, L. Fan, H. Babaei, Thermal conductivity enhancement of nanostructure-based colloidal suspensions utilized as phase change materials for thermal energy storage: a review, Renew. Sust. Energ. Rev. 24 (2013) 418–444.
[126] M. Mehrali, S.T. Latibari, M. Mehrali, H.S.C. Metselaar, M. Silakhori, Shape stabilized phase change materials with high thermal conductivity based on paraffin/graphene oxide composite, Energy Convers. Manag. 67 (2013) 275–282.
[127] A. Zabalegui, D. Lokapur, H. Lee, Nanofluid PCMs for thermal energy storage: latent heat reduction mechanisms and a numerical study of effective thermal storage performance, Int. J. Heat Mass Transf. 78 (2014) 1145–1154.
[128] K.V. Modi, H.K. Jani, I.D. Gamit, Impact of orientation and water depth on producti- vity of single-basin dual-slope solar still with Al_2O_3 and CuO nanoparticles, J. Therm. Anal. Calorim. 143 (2021) 899–913.
[129] H.H. Al-Kayiem, S.C. Lin, Performance evaluation of a solar water heater integrated with a PCM nanocomposite TES at various inclinations, Sol. Energy 109 (2014) 82–92.
[130] O. Bait, M. Si-Ameur, Tubular solar–energy collector integration: performance enhancement of classical distillation unit, Energy 141 (2017) 818–838.
[131] S.W. Sharshir, G. Peng, A.H. Elsheikh, E.M.A. Edreis, M.A. Eltawil, T. Abdelhamid, et al., Energy and exergy analysis of solar stills with micro/nano particles: a compara- tive study, Energy Convers. Manag. 177 (2018) 363–375.
[132] M.T. Chaichan, H.A. Kazem, Single slope solar distillator productivity improvement using phase change material and Al_2O_3 nanoparticle, Sol. Energy 164 (2018) 370–381.
[133] M.R. Safaei, H.R. Goshayeshi, I. Chaer, Solar still efficiency enhancement by using graphene oxide/paraffin nano–PCM, Energies 12 (10) (2019) 1–13.
[134] D.D.W. Rufuss, L. Suganthi, S. Iniyan, P.A. Davies, Effects of nanoparticle-enhanced phase change material (NPCM) on solar still productivity, J. Clean. Prod. 192 (2018) 9–29.
[135] G. Rajasekhar, M. Eswaramoorthy, Performance evaluation on solar still integrated with nano–composite phase change materials, Appl. Sol. Energy 51 (1) (2015) 15–21.

[136] V.K. Methre, M. Eswaramoorthy, Exergy analysis of the solar still integrated nano composite phase change materials, Appl. Sol. Energy 51 (2) (2015) 99–106.
[137] A.E. Kabeel, G.B. Abdelaziz, E.M.S. El-Said, Experimental investigation of a solar still with composite material heat storage: energy, exergy and economic analysis, J. Clean. Prod. 231 (2019) 21–34.
[138] A.E. Kabeel, M. Abdelgaied, A. Eisa, Effect of graphite mass concentrations in a mixture of graphite nanoparticles and paraffin wax as hybrid storage materials on performances of solar still, Renew. Energy 132 (2019) 119–128.
[139] A. Kandeal, N.M.E. Shafai, M.R. Abdo, A.K. Thakur, I.M.E. Mehasseb, I. Maher, M. Rashad, A.E. Kabeel, N. Yang, S.W. Sharshir, Improved thermo-economic performance of solar desalination via copper chips, nanofluid, and nano-based phase change material, Sol. Energy 224 (2021) 1313–1325.
[140] A.J. Chamkha, D.D.W. Rufuss, A.E. Kabeel, R. Sathyamurthy, M. Abdelgaid, A. M. Manokar, B. Madhu, Augmenting the potable water produced from single slope solar still using CNT-doped paraffin wax as energy storage: an experimental approach, J. Braz. Soc. Mech. Sci. Eng. 42 (2020) 625.
[141] H.R. Goshayeshi, I. Chaer, M. Yebiyo, H.F. Oztop, Experimental investigation on semicircular, triangular and rectangular shaped absorber of solar still with nano-based PCM, J. Therm. Anal. Calorim. 147 (2022) 3427–3439.
[142] A.S. Abdullah, F.A. Essa, H.B. Bacha, Z.M. Omara, Improving the trays solar still performance using reflectors and phase change material with nanoparticles, J. Energy Storage 31 (2020) 101744.
[143] M.M. Thalib, A.M. Manokar, F.A. Essa, N. Vasimalai, R. Sathyamurthy, F.P. C. Garcia Marquez, Comparative study of tubular solar stills with phase change material and nano-enhanced phase change material, Energies 13 (2020) 3989.
[144] A.E. Kabeel, R. Sathyamurthy, A.M. Manokar, S.W. Sharshir, F.A. Essa, A. H. Elshiekh, Experimental study on tubular solar still using graphene oxide nanoparticles in phase change material (NPCM's) for fresh water production, J. Energy Storage 28 (2020) 101–204.
[145] S. Shoeibi, H. Kargarsharifabad, N. Rahbar, Effects of nano-enhanced phase change material and nano-coated on the performance of solar stills, J. Energy Storage 42 (2021) 103061.
[146] F. Selimefendigil, C. Sirin, H.F. Oztop, Experimental analysis of combined utilization of CuO nanoparticles in latent heat storage unit and absorber coating in a single-slope solar desalination system, Sol. Energy 233 (2022) 278–286.
[147] S.S. Tuly, M.S. Islam, R. Hassan, B.K. Das, M.R.I. Sarker, Investigation of a modified double slope solar still integrated with nanoparticle mixed phase change materials: energy, exergy, exergo-economic, environmental, and sustainability analyses, Case Stud. Therm. Eng. 37 (2022) 102256.
[148] M. Elashmawy, M. Alhadri, M.M.Z. Ahmed, Enhancing tubular solar still performance using novel PCM-tubes, Desalination 500 (2021) 114880.
[149] D.D.W. Rufuss, S. Arulvel, V.A. Kumar, P.A. Davies, T. Arunkumar, et al., Combined effects of composite thermal energy storage and magnetic field to enhance productivity in solar desalination, Renew. Energy 181 (2022) 219–234.
[150] S. Shoeibi, H. Kargarsharifabad, N. Rahbar, G. Ahmadi, M.R. Safaei, Performance evaluation of a solar still using hybrid nanofluid glass cooling-CFD simulation and environmental analysis, Sustain Energy Technol Assess 49 (2022) 101728.
[151] B. Ellappan, V. Madhavan, Experimental investigation and economically productivity improvement analysis on single slope solar still from Ricinus Communis leaves with Ag nanoparticles doped phase change materials, Sol. Energy 261 (2023) 96–111.
[152] M. Almeshaal, S. Shanmugan, Improving thermal efficiency of solar stills: bioactive nano-PCM and Cramer's rule analysis, Sep. Purif. Technol. 343 (2024) 127119.
[153] K. Ganesan, Hybrid PV/T active two compartment solar still for improving performance of solar PV and enhance the productivity of solar still, Energy Sources A: Recovery Util. Environ. Eff. 45 (2) (2023) 5885–5904.
[154] S. Dhivagar, S. Shoeibi, H. Kargarsharifabad, M.H. Ahmadi, M. Sharifpur, Performance enhancement of a solar still using magnetic powder as an energy storage medium-exergy and environmental analysis, Energy Sci. Eng. 10 (8) (2022) 3154–3166.
[155] R. Dhivagar, S. Shoeibi, H. Kargarsharifabad, M. Sadi, A. Arabkoohsar, M. Khiadani, Performance analysis of solar desalination using crushed granite stone as an energy storage material and the integration of solar district heating, Energy Sources A: Recovery Util. Environ. Eff. 46 (1) (2024) 1370–1388.
[156] S.S.A. Toosi, H.R. Goshayeshi, I. Zahmatkesh, V. Nejati, Experimental assessment of new designed stepped solar still with Fe_3O_4 + graphene oxide + paraffin as nanofluid under constant magnetic field, J. Energy Storage 62 (2023) 106795.
[157] V.S. Vigneswaran, G. Kumaresan, B.V. Dinakar, K. Karthick Kamal, R. Velra, Augmenting the productivity of solar still using multiple PCMs as heat energy storage, J. Energy storage 26 (2019) 101019.
[158] R. Aggrawal, K.D.P. Singh, Performance evaluation of double slope solar still augmented with binary eutectic phase change material and steel wool fibre, Sustain Energy Technol Assess 48 (2021) 101597.
[159] R. Aggrawal, K.D.P. Singh, Experimental investigation and computational modelling of double slope solar still integrated with eutectic phase change material, J. Energy Storage 52 (2022) 104802.

Computing
https://doi.org/10.1007/s00607-024-01325-7

REGULAR PAPER



Efficient hashing technique for malicious profile detection at hypervisor environment

Anumukonda Naga Seshu Kumar[1] · Rajesh Kumar Yadav[1] · Nallanthighal Srinivasa Raghava[2]

Received: 27 November 2023 / Accepted: 10 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2024

Abstract

Attack detection in cyber security systems is one of the complex tasks which require domain specific knowledge and cognitive intelligence to detect novel and unknown attacks from large scale network data. This research explores how the network operations and network security affects the detection of unknown attacks in network systems. A hash based profile matching technique is presented in this paper for attack detection. The main objective of this work is to detect unknown attacks using a profile matching approach in Hypervisors. Hypervisors are characterized by their versatile nature since they allow the utilization of available system resources. The virtual machines (VMs) in the hypervisors are not dependent on the host hardware and as a result, hypervisors are considered advantageous. In addition, hypervisors have direct access to the hardware resources such as memory, storage and processors. However, hypervisors are more susceptible to the security threats which attack each and every VM. A SHA3-512 hashing algorithm used for generating hash values in hypervisor and the proposed model is used to verify whether the profile is malicious or benign. The performance of the hashbased profile matching technique is compared with traditional hash techniques namely SHA-256 and MD5 algorithm. Results show that the proposed SHA3-512 algorithm achieves a phenomenal performance in terms of phenomenal accuracy and zero false positive rates. Simulation results also show that the computation time required by Sha3-512 algorithm is lower compared to SHA-256 and MD5 algorithms. The performance analysis validates that the hash based approach achieves reliable performance for attack detection. The effectiveness of the hashing technique was determined using three different evaluation metrics namely attack DR, FPR, and computational time. Simulation results show that the existing SHA3- 512 algorithm detection rate of 97.24% with zero false positive rate and faster computational time compared to SHA 256 and MD5 algorithms.

Keywords Attack detection · Hash-based profile matching · Cryptographic hash algorithms · Sha3-512 hashing algorithm · Detection accuracy · False positive rate

Extended author information available on the last page of the article

Published online: 19 July 2024

Mathematics Subject Classification 68M99

1 Introduction

Cyber security and security of network systems are the fastest growing paradigms in recent times. Injection of cyber-attacks is an attempt to disrupt the operations of network systems and make them unavailable to the users [4]. Most of the time, the attacks are novel and it is highly challenging to detect and defend such attacks because of its unknown nature [15]. Though several techniques are available for defending unknown attacks in the systems, these attacks keep evolving and transforming themselves which makes it complicated to detect them [16, 25]. With the evolution of advanced network systems and technological enhancements, the attackers are finding different ways to introduce unknown attacks into the network system. In general, there are two forms of attacks, one that disrupts the network operation and another one that crashes the system completely.

Conventional attack detection techniques such as signature based systems [9] and intrusion detection systems [3] have been used extensively in several approaches. However, these attacks are limited to the detection of known or predefined attacks. Few researchers have also suggested the application of cryptography based encryption techniques along with intrusion detection techniques [5, 10, 26]. Cryptography based techniques have yielded superior results in terms of protecting the security of network data [18, 28]. Cryptographic hashing is another technique which has gained huge popularity in recent times for attack detection [7, 22]. Cryptographic hashing is different from conventional cryptography based data encryption techniques. Encryption defends passive attacks, while cryptographic hashing algorithms are used for creating a digital fingerprint of the data which helps in preventing active attacks in the network. Each digital fingerprint is unique and hence it can prevent falsification of attacks and thereby enable secure authentication of data [17].

Hence, Secure Hashing Algorithms (SHAs) are employed for ensuring the security and privacy of network systems [29]. Hashing algorithms are considered to be more appropriate for securing the network applications which contain sensitive and confidential information. One of the prominent features of SHAs is that even a smaller variation in the input results in the significant variation in the output [14, 24]. According to NIST, the SHA algorithms are categorized into different hash algorithms namely SHA- 0, SHA-1, SHA- 2, and SHA- 3. The SHA- 0 algorithms was the initial version of the 160- bit hashing algorithm which was introduced in 1993. As a result of the continuous transformation of SHA, the improved version of SHA-1 was developed. The SHA- 1 algorithm consists of a 160- bit function and is similar to the fundamental MD 5 algorithm. The SHA -1 algorithm was designed to be an integral part of the Digital Signature Algorithms (DSA) [17]. Most of the limitations of cryptographic methods were observed after the implementation of SHA- 1 algorithm and hence this algorithm was restricted from being implemented after 2010, for securing the network information. The drawbacks of SHA- 1 algorithm were resolved in another version of

SHA algorithm namely SHA- 2, which is the integrated form of peer hash functions such as SHA- 256 and SHA- 512 which has varying block sizes.

Few other modified versions of the SHA are introduced with different specifications as shown in Table 1 [20, 30]. The SHA-2 hash functions are one of the widely used hashing techniques compared to other algorithms. However, the development of SHA- 3 hash function has improved the attributes of SHA- 2 algorithms [8].

The variants of SHA3 are discussed in Table 1. This research aims to address these challenges by presenting an efficient hash-based profile matching technique for detecting unknown attacks, utilizing the SHA-3-512 algorithm. This algorithm generates unique hashes for each profile, enhancing the detection of malicious activities in network systems. The novel SHA-3 can be employed in various tasks that require robust security such as banking, and financial applications, biometric techniques for authentication etc.

The contributions of this research are as follows:

This paper presents an efficient hash-based profile matching technique for detecting unknown attacks in hypervisors.
A Sha3- 512 hashing algorithm is implemented in this work, which generates a unique hash for each and every profile.
The proposed hash-based profile matching model is provided with new inputs to verify whether the profile is benign or malicious.
The proposed model works based on the profile length and profile data for detecting the attacks. The efficacy of the profile matching model is evaluated based on the detection ratio (DR) and false positive rate (FPR).
A comparative analysis is conducted wherein the potentiality of the Sha3- 512 algorithm is validated with respect to SHA-256 and MD5 algorithms.

The article is further structured as follows: Sect. 2 discusses the existing works related to cryptography based attack detection. Section 3 discusses the hash-based profile matching technique for detection of unknown attacks and how the proposed system is developed. Section 4 outlines the results of the proposed methodology and Sect. 5 presents the conclusion.

Table 1 The variants of SHA3 algorithm and its specification

SHA 3 Variant	Output size (bits)	Block size (bit rate r) (bits)	Capacity c (bits)	Security
SHA3 -224	224	1152	448	112
SHA3 -256	256	1088	512	128
SHA3 -384	384	832	768	192
SHA3 -512	512	576	1024	256

2 Related work

Various literary works have been published for identifying security attacks in cyber network systems [6, 11, 13]. A hash based contextual signature is used for detecting unauthorized intrusions in the network [21]. The proposed approach intends to overcome the drawbacks of conventional IDSs using hash based contextual signature which integrates the intrusion detection signatures with hash functions. The contextual signatures are used to strengthen the security of the contextual information and its ability to filter out redundant alarms. The performance of the IDS was tested considering the probability of selecting relevant hash functions and its adaptability in a real time network environment. Results of the experimental analysis show that the proposed approach is highly effective in minimizing the non-critical alarms. Sharma [27] reviewed different encryption techniques. The significance of cryptography in information security is proposed by [12]. The study validates that cryptography based techniques are highly efficient in protecting confidential data from unauthorized users. The security of data is ensured using encryption and decryption which also ensures that the privacy of communication is maintained. This research states that without cryptographic techniques the sensitive information can be easily accessed by the intruders and cryptographic methods encrypt the data into an unreadable format and thereby ensures that the data is not accessible to unverified users. This research also discusses recent advancements in the field of cryptography, state of art of cryptographic techniques and different types of cryptography. Although cryptography ensures high level security of confidential information, it cannot be ensured that the data is completely protected from unauthorized access. To overcome this, the cryptography techniques are improvised by using hashing algorithms. The work presented by Maetouq et al. [19] compared different hash function algorithms for attack detection. Different hashing algorithms are studied to assure authentication and privacy of the information using MD5, SHA and RIPEMD. Different algorithms are analyzed and their limitations are discussed. This work suggests that the performance of the existing hash algorithms is computationally intensive.

3 Proposed research methodology

A Sha3-512 hashing algorithm is adopted in this paper. The proposed approach is based on profile matching and in this hashing technique, the hash values which are in the form of strings are converted into hash structure with a key value. The data structure of the hash map is directly mapped into the relevant data using a given key for classifying the attacks. If no matching key is found, the hash table is updated with the newly generated key and the updated table is mapped to each location. This process is continued on a regular basis until all the matching keys are identified. The stages involved in the implementation of the Sha3-512 hashing technique are described in Fig. 1.

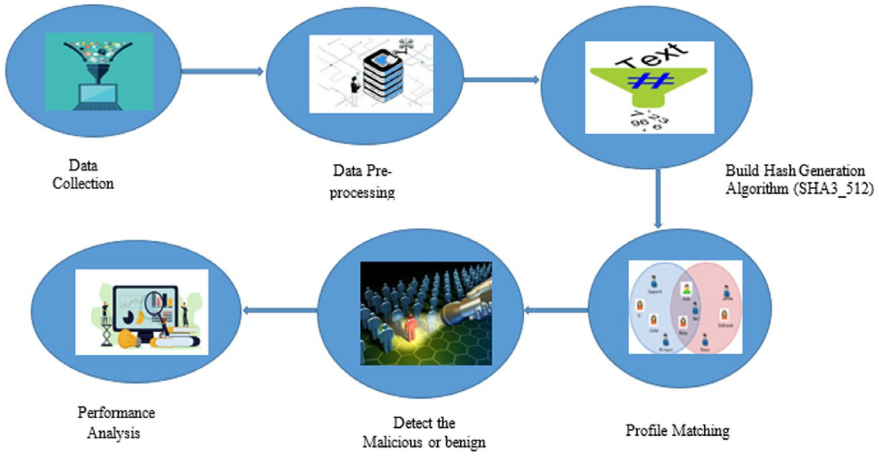


Fig. 1 Workflow of the proposed approach

3.1 Data collection and preparation

The data for the experimental analysis is collected from the HYPERVISORS. A hypervisor is software which creates and executes virtual machines (VMs). It distinguishes the hypervisor operating systems with VMs and thereby allows the creation and execution of VMs. Hypervisors when used as physical hardware are called hosts and the VMs that employ the resources of hosts are defined as guests. The hypervisor software treats different elements of computer systems such as memory, processing units and storage units as a pool which can be shifted from one VM to another.

The data collected from the Hypervisor cloud dataset [1] consists of three different network traffic scenarios namely, internal, external, and local network traffic. The external data flow is between the outside machine and the virtual machine. Meanwhile, the internal traffic is defined as the data flow between one hypervisor to another (this traffic is known as hypervisor traffic). On the other hand, the local traffic happens between two instances on the same hypervisor and this traffic is called instance-instance traffic. In this research, two types of data are collected from the network traffic: (i) network traffic data without any payload on the instances and the hypervisor (ii) complete network traffic focused only on the hypervisors. The dataset consists of both malicious and benign network data packets. The total numbers of data packets in the datasets are 22,372,418, out of which 22,356,769 packets are benign and 15,649 are malicious [2].

3.2 Data pre-processing

Data pre-processing is the primary and most important step in most of the attack detection tasks. It is highly significant in enhancing the accuracy of attack detection techniques. In general, the data is pre-processed to eliminate redundant data from the input data. Pre-processing is carried out to remove the external noise, to

prevent the influence of uncertainties, removal of null values, and filtering of missing data. The cyber security dataset includes attack related features which are used for the classification of unknown attacks. From the available set of attributes, only the relevant features are considered for the detection process based on the respective correlation.

3.3 Hash generation

In the proposed research, a hash value is generated for each profile. As discussed previously, a Sha3- 512 hashing is used for hash generation. The hash generation process is executed using different input features such as (i) source Ip, (ii) dest Ip, (iii) mac_address_sourceIp (iv) mac_address_dest Ip, (v) seq Number (vi) ack Number. Further, the inputs are given to the Sha3- 512 hashing, based on which the algorithm generates a unique hash value for each and every profile.

The process of hashing involves the application of hash functions to the data items using a hashing key to generate a hash value. Hashing algorithms use values of larger range defining all possible data strings. These values are mapped to the smaller set of values such as an18-bit number. There are two important applications involved in the hashing process i.e., improving the speed of data retrieval and validation of data. Hashing algorithms make use of hash values to perform complex calculations and generate an output with a fixed length string. The main intention behind using a hashing algorithm in this research is its non-reversible nature. In general, hash values cannot be reversed and obtaining the result of hash value of a file from the hashing algorithms does not ensure the reconstruction of file's information. However, the hash values can be used to determine whether the profile of two files are identical or not without even analyzing the information of the files. Hence, the concept of utilizing hashing algorithms for profile matching is unique and is applicable to the entire hash functions. If two files can generate the same information then this process is termed as collision and in such cases, it is not possible to use the hash values for matching the profiles. To avoid collision, this research intends to use SHA3- 512 hashing algorithm.

3.4 Proposed secure hash algorithm (SHA3- 512)

As discussed previously, hashing algorithms suffer from collision attacks which affect the performance of the algorithms. In this case, the SHA- 3 hashing is regarded as the advanced algorithm. Cryptographic hash secures the data by generating a unique hash value for each data sample. For comparing two or more data samples, it is suggested to perform hashing first. Even the slightest change in the algorithm will generate different hash values. This attribute makes the SHA algorithms more suitable for verifying the integrity of any information along with other technologies such as digital signatures and anti-tampering. The process involved in the SHA3-512 technique is illustrated in Fig. 2.

A cryptographic hash function known as the SHA3-512 algorithm generates a fixed-size output of 512 bits from an input message of arbitrary length. The

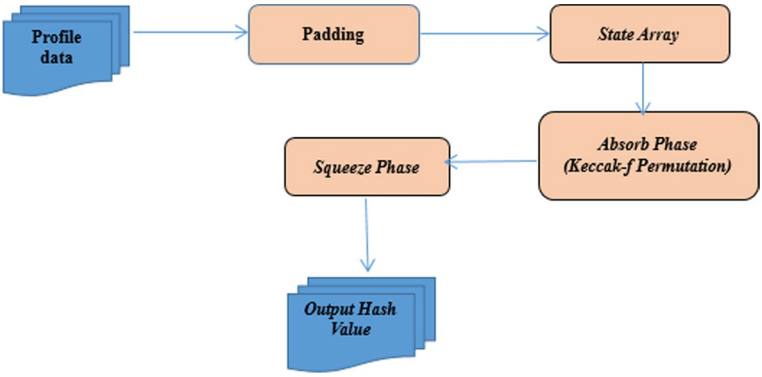


Fig. 2 Process of the SHA3-512 algorithm

algorithm is based on the Keccak sponge function, which is a type of hash function that can be used to create hash functions with arbitrary output lengths. When using the SHA3-512 algorithm, the input message is first padded to make it a multiple of the block size, which is 72 bytes in this case. After that, each block of the padded message has been divided up and processed by the Keccak sponge function.

3.4.1 Keccak sponge function

The KECCAK is the final version of the SHA-3 algorithm. The sponge function used in this research is designed using two stages namely the absorbing stage and the squeezing stage. This is responsible for mixing and transforming the internal state of the hash function during the computation of the hash value. The permutation operates on a state array, which consists of a grid of lanes. The structure of the permutation is characterized by a few key features:

3.4.1.1 Padding The input message is padded in the first unit to make sure it is the right length. The padding phase creates padded messages that are length multiples of the bitrate r in order to strengthen and enhance Keccak's defenses from length extension attacks. The padding process involves adding a series of bits to the end of the input message to bring it up to the required block size. The first step in the padding procedure is to finish the message with a single "1" bit. When the message length is one bit short than a multiple of the block size, a string of "0" bits are appended in succession. Using a fixed-length encoding scheme, the original message's length is then added to the end of each padded message. The goal of the padding procedure is to guarantee that the input message meets the Keccak sponge function's requirement that it be a multiple of the block size.

3.4.1.2 State array The state is made up of a matrix of 5×5 64-bit words. Blocks of the padded message that are individually the same size as the state array are then created. For SHA-3–512, each block is 72 bytes (576 bits).The state 'S' is considered as

a 3-D array with x, y, and z as three coordinates. The 3-D state is represented as 1-D array state which is denoted by 'A' and is determined as shown in Eq. 1:

$$A[x, y, z] = S[w(5y + x) + z] \text{ Where } 0 \leq x \leq 4,\ 0 \leq y \leq 4, \text{ and } 0 \leq z \leq (w - 1) \tag{1}$$

Here 'w' is a parameter representing the width of the state array, which is related to the number of bits in each lane.The array state in this research incorporates lanes of dimension 5×5 with each lane consisting of 64 bits.In other words, the array state consists of 64 segments wherein each segment is made up of 25 bits. For the proposed SHA3- 512 algorithm, $w = 64$ and the bit rate is equal to 1600. In the initial process, the hash function converts the input information with an arbitrary length of fixed size.

3.4.1.3 Absorbing phase Every message block gets XOR-ed with the initial r bits of the state during the absorption phase, which is followed by the implementation of the permutation f. The Keccak permutation, a series of mathematical processes intended to offer a maximum level of security, is used in SHA-3-512. The state array is mixed and altered iteratively using the Keccak permutation. Up untill all message blocks have been handled, this procedure is continued.

3.4.1.4 Squeezing phase After the absorbing phase is complete, the state array contains a fixed amount of information derived from the input data. In the squeezing phase, this information is extracted to produce the final hash value. The squeezing phase may involve multiple iterations of the Keccak permutation to obtain the desired hash length of 512 bits.The first r bits of the state are returned as (part of) the output at each iteration, and the current state is then subjected to the permutation*f*.

Figure 3 depicts the block diagram of the sponge construction. Keccak-based hash functions, i.e., SHA-3, use a sponge construction. This means they absorb input data into the state array and then squeeze out the hash value. An input called state is given to the algorithm as a matrix. The state is made up of a three-dimensional 5×5

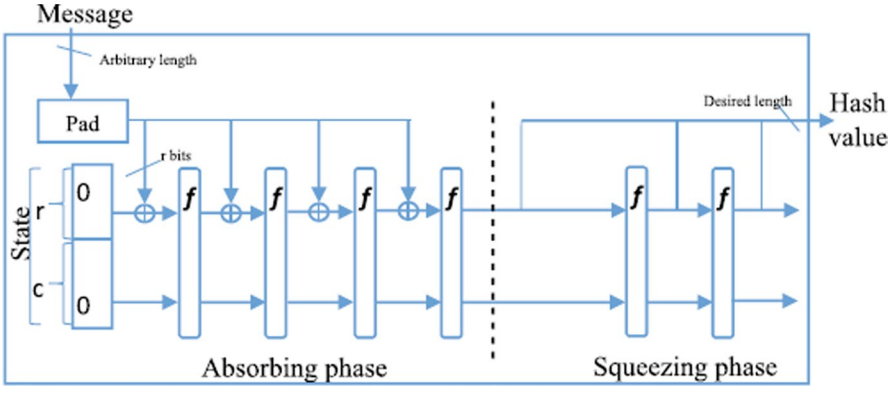


Fig. 3 Keccak sponge function

word-size table with a length of 1600 bits, where word-size $\in \{1, 2, 4, 8, 16, 32, 64\}$. Two parameters are used by the method to form the sponge, based on the required output length. The two input parameters that affect the security level that can be achieved are the bit rate with r bits as well as the capacity with c bits. The input message has been padding and split into r-bit blocks. The primary processing phase, function f, has 24 cycles with processes. A state with width is affected by the function f, whereby r denotes the outer state and c the inner state. The Keccak permutation operates in round functions, and each round applies a series of bitwise operations to the state array. For SHA-3-512, the Keccak-f permutation operates for 24 rounds. These operations include:

The Keccack function KECCACK-f (1600) contains 24 permutation rounds, and each round is made up of five successive phases designated in the state matrix A as theta, rho, pi, chi, and iota, denoted as θ, ρ, π, χ, τ respectively. The algorithm receives the state matrix A as well as the round constant as inputs in every iteration Round[b] (A, RC), producing a hash code output of size b for the following round. The following formulas illustrate the information flow through these five KECCACK-f phases (Fig. 4).

Step theta (θ)

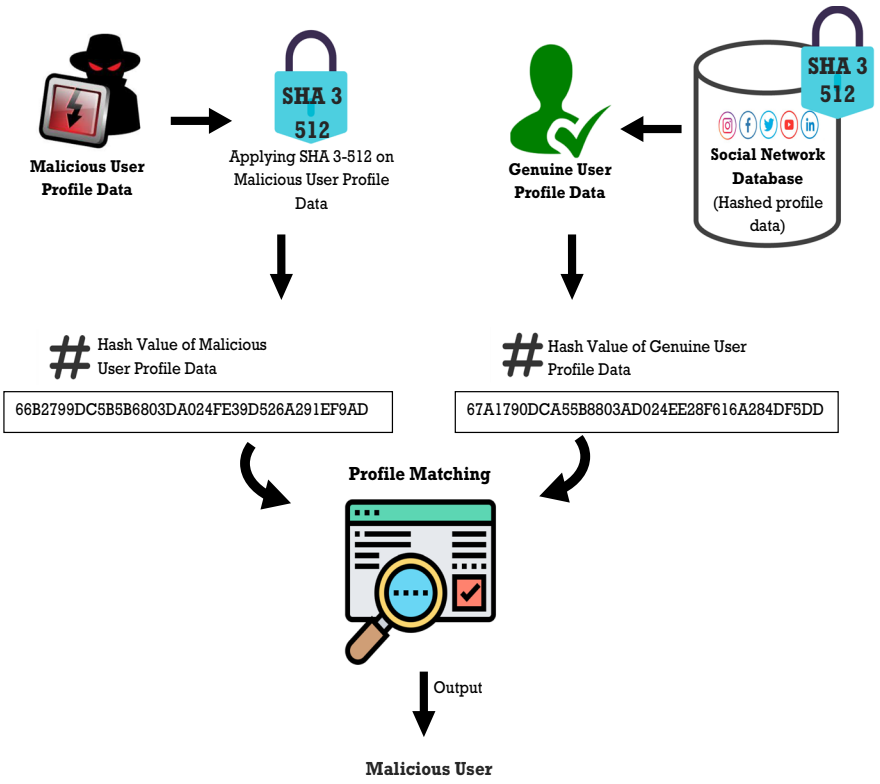


Fig. 4 Profile matching model

Equation (2) demonstrates that theta is composed of a bitwise XOR, a rotation of one position, and a parity calculation.

$$\begin{aligned} C[x] &= A[x,0] \oplus A[x,1] \oplus A[x,2] \oplus A[x,3] \oplus A[x,4] \quad 0 \le x \le 4 \\ D[x] &= C[x-1] \oplus Rot([x+1],1) \quad 0 \le x \le 4 \\ A[x,y] &= A[x,y] \oplus D[x] \quad 0 \le x,y \le 4 \end{aligned} \tag{2}$$

Step Rho (ρ) and pi (π)

In Eq. (3), pi is a permutation, and Rho is a rotation with an offset that is dependent upon the word position.

$$B[y, 2x+3y] = Rot(A[x,y], r[x,y]),\ 0 \le x,y \le 4 \tag{3}$$

Step Chi (ρ)

Equation (4)'s Chi is made up of bitwise XOR, NOT, and AND gates.

$$A[x,y] = B[x,y] \oplus ((NOT\, B[x+1,y])\, AND\, (B[x+2y]) \qquad 0 \le x,y \le 4 \tag{4}$$

Step iota (τ)

Iota in Eq. (5) is a constant round addition, to sum things up.

$$A[0,0] = A[0,0] \oplus RC[i] \tag{5}$$

Here, *RC[i]* represents the round constant for round i. A round is finished once these five steps are finished. The bits that must be processed are completed by the Padding function. B[x, y], C[x], and D[x] are the intermediary variables.Rot () denotes the binary cyclic shift operation. Once the hash generation process is completed, the model is tested for performance evaluation. During testing, the new inputs to the hash based profile matching model are generated and based on the new inputs; the model will check whether the profile is benign or malicious (Fig. 5).

3.5 Profile matching

After hash generation, the hashed profile data is stored in the database. When a new input profile is obtained, same hashing function is performed on the new input using SHA 3-512 algorithm. This model compares the hash value of each new input with the hash value of each stored input in the database. The user is referred to as "Benign" if both the hash values of the stored input and new input match, otherwise the user is referred to as "Malicious" if they both differ with each other. If both the hash values of stored input and new input match, then the user is “Benign” user, otherwise the user is referred to as "Malicious" as if both the hash values differ with each other.

The steps involved in SHA3-512 Hashing algorithm are explained as follows:

Step 1 Obtain the user profile data from the social network database.

Step 2 Perform hash function on the obtained user profile data using the SHA-3–512 hashing algorithm.

Step 3 Store the hashed profile data of the user in the database, associating it with the user's profile features (e.g., user ID).

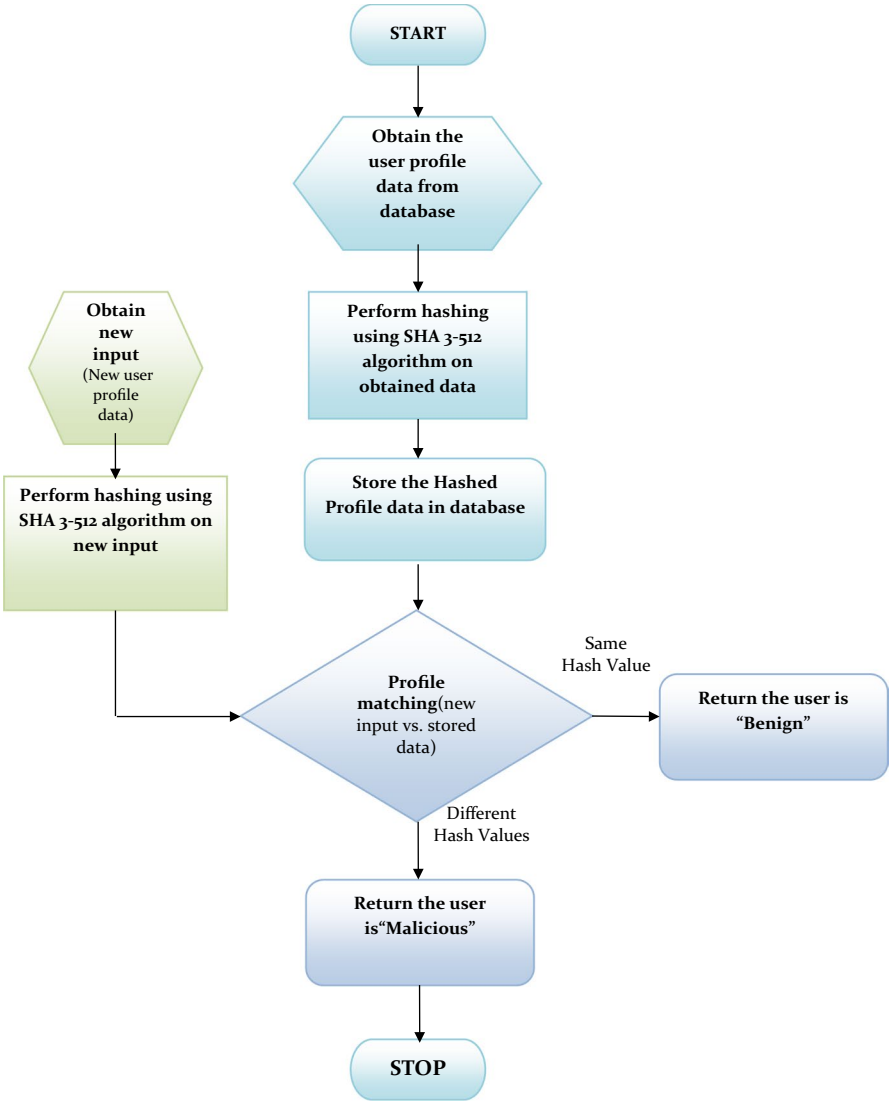


Fig. 5 Flowchart of proposed methodology

Step 4 Get new user input profile data to verify whether the user is malicious or benign.

Step 5 Repeat Step 2 for the new input data and hash it using the same SHA-3–512 hashing algorithm.

Step 6 Execute profile matching to verify the user's status:

(i) Compare the hash value of the new input with the hash value of the stored input in the database.

(ii) If the hash values match, return "Benign" (indicating the user is benign).
(iii) If the hash values do not match, return "Malicious" (indicating the user is malicious).

Step 7 Return whether the user is “malicious” or “benign”The overall flow of the proposed work is represented as below;

Once the testing of the hashing process is completed, the performance is evaluated based on the DR and FPR. The outcome of the simulation analysis is discussed in the next section.

4 Experimental results

A brief overview of the simulation of the SHA3- 512 hashing algorithm and its attack detection ability in terms of three performance evaluation metrics namely DR, FPR and computational time. Two simulations were performed using the same executable file without employing any optimization levels. The performance of SHA3- 512 algorithm was also analyzed without using any compiler optimization. In general compiler optimizations are employed for improving the performance of simulated models and several researchers have determined the influence of code optimization on different aspects of simulation such as reliability of the model and energy consumption. In addition, it was observed from the simulation analysis that the optimization of the compiler minimizes excess consumption of system energy by minimizing the number of instructions of the machine required for executing any operation [23]. Furthermore, the minimization of the number of computations affects the execution of software programs and thereby reduces the hardware errors. It was also inferred from the simulation analysis that the compiler optimizations have a negligible effect on the error sensitivity of different hashing algorithms.

4.1 Performance evaluation

The performance of the hashing technique was determined in terms of DR, FPR, and computational time. The results of the simulation analysis are compared with other two conventional hashing algorithms namely SHA 256 and MD5 (Message Digest) algorithm and other models such as Snort (ET rule set), Snort (Custom rule set) and a hypervisor based IDS.

The SHA 256 is one of the cryptographic hash functions which is used to compare two sets of data. The SHA 256 algorithms can identify even small variations in the data and it generates a different hash value which indicates the difference in the data. The algorithm generates a unique, and fixed size 256-bit (32-byte) hash which makes it appropriate for validating the integrity of the data by identifying the tampering of the digital signatures. On the other hand MD5 algorithm is a one-way hash function which makes it practically impossible to retrieve the original data from its hash value. The MD5 algorithm generates a unique 128-bit hash value for each input. However, MD5 is not used extensively in cryptographic applications because

Table 2 Comparative analysis

Algorithms	Detection rate (DR)	False positive Rate (FPR)	Computational time (Sec)
SHA3 -512	97.240	0.0	30.261
SHA 256	76.155	0.063	29.038
MD5	87.934	0.536	25.141
Snort (ET rule set)	10.5	0.002	26.380
Snort(Custom rule set)	41.23	0.0065	24.754
Hypervisor-based IDS	96.11	7.3	27.148

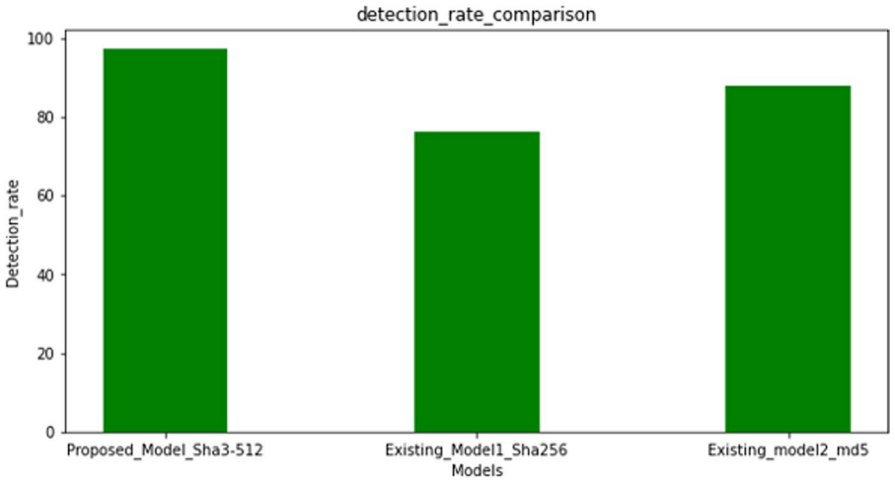


Fig. 6 Detection rate of different hashing models

of its vulnerabilities and SHA-256 is mostly preferred as an alternative. Snort is one of the widely used open source IDS which provides network traffic information in real time. One of the main functionality of Snort emerging threats rule set (Snort (ET rule set)) is its updated set of rules which improves the intrusion detection capability of Snort. The Snort(Custom rule set) helps the researchers to customize the IDS based on their requirements. This enables the model to detect threats precisely and respond to the threat. The custom rule set provides proactive security measures and strengthens the security of the organization in terms of addressing their concerns. A hypervisor-based IDS operates at the hypervisor level by exploiting the advantages of the virtualization technology for monitoring and detecting malicious activities in the virtualized environments. It is mainly designed to provide a comprehensive analysis of the security threats and protect VMs running on hypervisors.The outcome of the comparative analysis is tabulated in Table 2.

The graphs depicting the DR, FPR and computational time are discussed in the following Figs. 6, 7, 8.

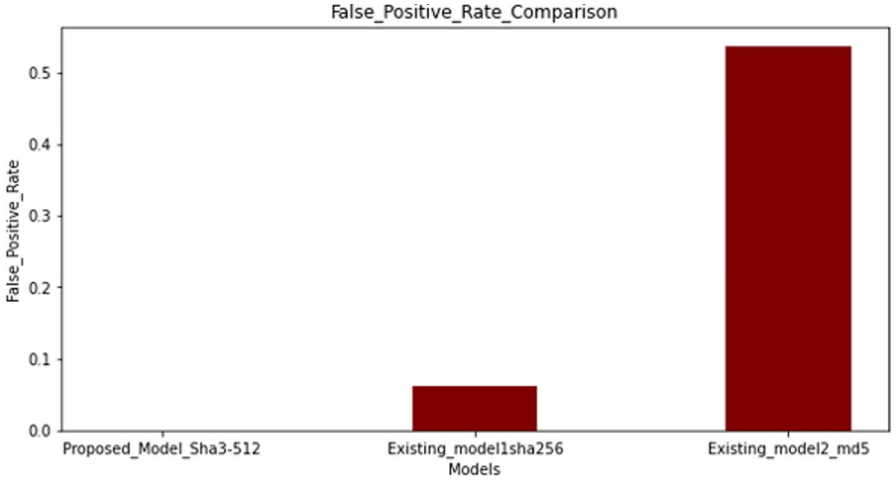


Fig. 7 FPR of different hashing models

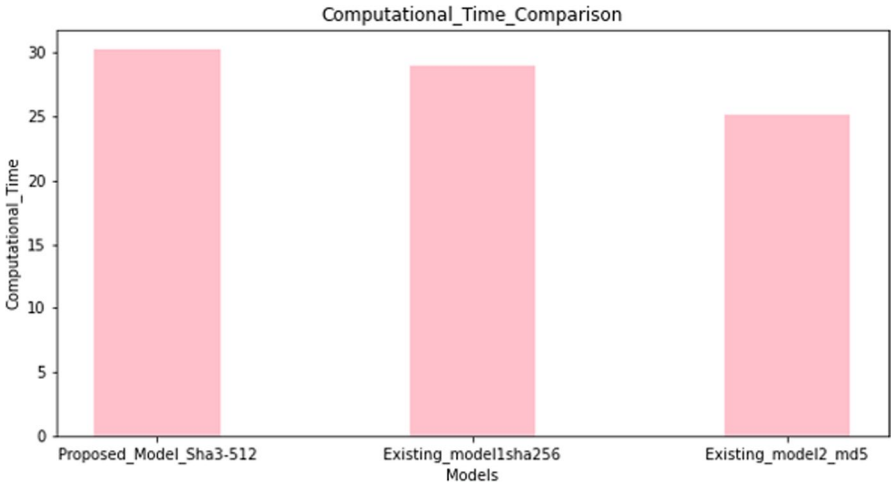


Fig. 8 Computational time taken by of different hashing models

It can be inferred from the results of the comparative analysis that the proposed SHA3- 512 hashing algorithm achieves a DR of 97.24% with zero FPR and faster computational time. The model takes only 30.26 s for executing the hashing process. On the other hand, the MD5 algorithm performs better compared to SHA 256 algorithms by achieving an attack detection rate of 87. 93% with a FPR of 0.53. However, the performance of SHA 256 algorithm is better with respect to FPR. It achieves a FPR of 0.06 which is very less compared to SHA 256 hashing algorithm. In terms of hierarchy, it can be inferred that the performance of the existing SHA3- 512 hashing algorithm is better compared to the other two algorithms.

The SHA-3-512 algorithm is inherently more computationally efficient. This cryptographic hashing method simplifies the detection process, reducing the computational time and resources needed. Our evaluation has demonstrated that the SHA-3-512 algorithm not only achieves superior detection rates and lower false positive rates but also completes the hashing process in a much shorter time frame, specifically in 30.26 s. The performance of the proposed approach is also compared in terms of detection accuracy and the obtained results are tabulated in Table 3. The existing results mentioned in Aldribi et al. [2].

The detection comparison is schematically in Fig. 6.

Correspondingly, the global detection results that define the performance of the existing hypervisor presented in Aldribi et al. [2] and proposed hypervisors are tabulated in Table 4.

As inferred from the results tabulated in Table 4, the proposed hypervisor based on the profile matching technique achieves excellent results compared to existing hypervisors (Fig. 9).

5 Conclusion

This paper presents an SHA3- 512 hashing based profile matching approach for detecting unknown attacks. The proposed model works based on the profile length and profile data. As observed from existing literary works, there are several types of hashing algorithms employed for ensuring the confidentiality and authentication of the information. Few hashing algorithms such as variants of SHA have proven their ability to provide security to the information. However, this research implemented SHA3- 512 algorithm to avoid the effect of collision. A comprehensive review of existing algorithms was conducted and it was observed that most of the existing algorithms are computationally intensive. In addition, the paper also reviews other hashing techniques such as AES, DES, which were discussed in the existing literary works. Based on the observations, this study concludes that the hash function is highly efficient and is capable of satisfying the requirements for attack detection such as data confidentiality, and authenticity. The effectiveness of the hashing technique was determined using three different evaluation metrics namely attack DR, FPR, and computational time. Simulation results show that the existing SHA3- 512 algorithm detection rate of 97.24% with zero false positive rate and faster computational time compared to SHA 256 and MD5 algorithms. Results validate the efficacy of the SHA3- 512 algorithms in terms of detecting unknown attacks compared to existing hypervisors and Snort techniques. While this study demonstrates the efficacy of the SHA3-512 hashing based profile matching approach for detecting unknown attacks in static environments, several limitations and opportunities for future work are identified. One notable limitation is the current focus on static environments, which may not fully capture the dynamic nature of modern networks, such as IoT environments where devices and threats evolve rapidly. To address this, future research could extend the SHA3-512 approach to dynamically adapting environments by developing algorithms that can effectively handle changing network conditions and evolving attack vectors in real-time. However, the evaluation primarily

Table 3 Comparison of detection accuracy

IDS	Day 1		Day 2		Day 3		Day 4		Overall	
	FPR%	DR%	FPR%	DR%	FPR%	DR%	FPR%	DR%	FPR%	DR%
Snort (ET ruleset)	0.00003	0.10	0.0001	9.84	0.043	8.8	0	17.5	0.002	10.5
Snort (Custom ruleset)	0.011	69.01	0.0047	31.04	0.003	30.52	0.0075	34.36	0.0065	41.23
Hyp-based	0.82	98.1	9.9	9.9	7.13	84.2	6.7	100	7.3	96.11
Hyp profile matching based	0	99.57	0.0	100.0	0.0	98.42	0.0	100.0	0.0	99.70

Table 4 Comparison of hypervisors

Phases	Existing Hypervisors				Proposed Hypervisor	
	Hypervisor A		Hypervisor B		Profile matching technique based Hypervisor	
	FPR%	DR%	FPR%	DR%	FPR%	DR%
Phase 1	10.42	94.77	7.34	96.11	0.0	99.70
Phase 2	13.23	96.70	5.66	95.94	0.0	97.68
Overall Performance	FPR = 7.56, DR = 96.23				0.01	98.35

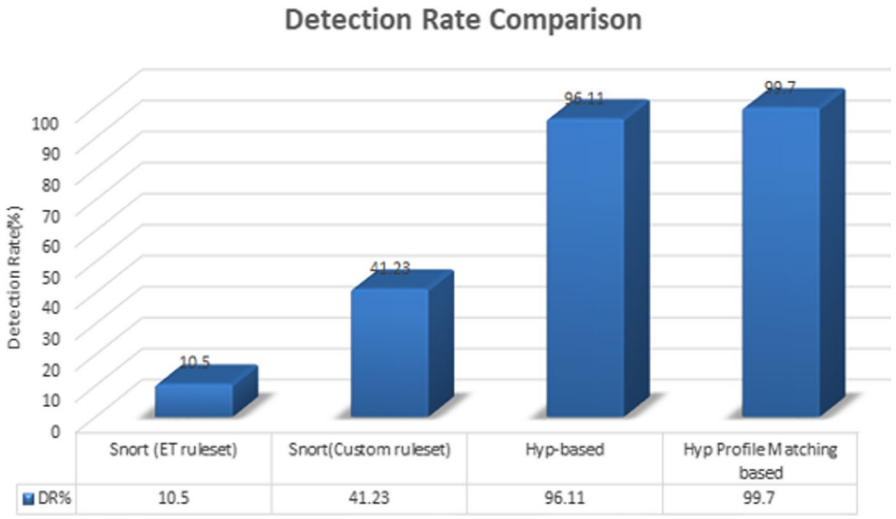


Fig. 9 Detection rate comparison

focuses on limited traditional metrics like DR, FPR, and computational time. Future evaluations could include more diverse attack scenarios, varying network conditions, and scalability tests to better reflect real-world deployment scenario.

Author contributions All authors read and approved the final manuscript.

Funding The authors declare that they have competing interests and funding.

Data availability Data sharing is not applicable to this article because of proprietary nature.

Code availability Code sharing is not applicable to this article because of proprietary nature.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Aldribi A, Traore I, Moa B (2018) Data sources and datasets for cloud intrusion detection modeling and evaluation. Cloud computing for optimization: foundations, applications, and challenges. Springer, Cham, pp 333–366
2. Aldribi A, Traoré I, Moa B, Nwamuo O (2020) Hypervisor-based cloud intrusion detection through online multivariate statistical change tracking. Comput Secur 88:101646
3. Alhakami W, Alharbi A, Bourouis S, Alroobaea R, Bouguila N (2019) Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection. IEEE Access 7:52181–52190
4. Ben-Asher N, Gonzalez C (2015) Effects of cyber security knowledge on attack detection. Comput Hum Behav 48:51–61
5. Canard S, Diop A, Kheir N, Paindavoine, M, Sabt M (2017). BlindIDS: market-compliant and privacy-friendly intrusion detection system over encrypted traffic. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security. pp. 561–574. https://doi.org/10.1145/3052973.3053013
6. Chen Y, Kar S, Moura JM (2016) Dynamic attack detection in cyber-physical systems with side initial state information. IEEE Trans Autom Control 62(9):4618–4624
7. Cui A, Li M, Qu G, Li H (2020) A guaranteed secure scan design based on test data obfuscation by cryptographic hash. IEEE Trans Comput Aided Des Integr Circuits Syst 39(12):4524–4536
8. Dworkin, MJ (2015) SHA-3 standard: permutation-based hash and extendable-output functions. https://doi.org/10.6028/NIST.FIPS.202
9. Einy S, Oz C, Navaei YD (2021) The anomaly-and signature-based IDS for network security using hybrid inference systems. Math Probl Eng. https://doi.org/10.1155/2021/6639714
10. Ferriyan A, Thamrin AH, Takeda K, Murai J (2021) Generating network intrusion detection dataset based on real and encrypted synthetic attack traffic. Appl Sci 11(17):7868
11. Gao L, Chen B, Yu L (2019) Fusion-based FDI attack detection in cyber-physical systems. IEEE Trans Circuits Syst II Express Briefs 67(8):1487–1491
12. Gençoğlu MT (2019) Importance of cryptography in information security. IOSR J Comput Eng 21(1):65–68
13. Giraldo J, Urbina D, Cardenas A, Valente J, Faisal M, Ruths J, Tippenhauer NO, Sandberg H, Candell R (2018) A survey of physics-based attack detection in cyber-physical systems. ACM Comput Surv (CSUR) 51(4):1–36
14. Harran M, Farrelly W, Curran K (2018) A method for verifying integrity & authenticating digital media. Appl Comput Inform 14(2):145–158
15. Huda S, Miah S, Hassan MM, Islam R, Yearwood J, Alrubaian M, Almogren A (2017) Defending unknown attacks on cyber-physical systems by semi-supervised approach and available unlabeled data. Inf Sci 379:211–228
16. Kamarudin MH, Maple C, Watson T, Safa NS (2017) A logitboost-based algorithm for detecting known and unknown web attacks. IEEE Access 5:26190–26200
17. Karmani M, Benhadjyoussef N, Hamdi B, Machhout M (2021) The SHA3-512 cryptographic hash algorithm analysis and implementation on the Leon3 processor. Int J Eng Trends Technol 69(6):71–78
18. Li J, Zhang Y, Ning J, Huang X, Poh GS, Wang D (2020) Attribute based encryption with privacy protection and accountability for CloudIoT. IEEE Trans Cloud Comput 10(2):762–773
19. Maetouq A, Daud SM, Ahmad NA, Maarop N, Sjarif NNA, Abas H (2018) Comparison of hash function algorithms against attacks: a review. Int J Adv Comput Sci Appl. https://doi.org/10.14569/IJACSA.2018.090813
20. Martino R, Cilardo A (2020) SHA-2 acceleration meeting the needs of emerging applications: a comparative survey. IEEE Access 8:28415–28436
21. Meng Y, Kwok LF (2014) Adaptive non-critical alarm reduction using hash-based contextual signatures in intrusion detection. Comput Commun 38:50–59
22. Mohindru V, Singh Y, Bhatt R (2020) Hybrid cryptography algorithm for securing wireless sensor networks from node clone attack. Recent Adv Electr Electron Eng (Former Recent Pat Electr Electron Eng) 13(2):251–259

23. Nazarian G, Strydis C, Gaydadjiev G (2011) Compatibility study of compile-time optimizations for power and reliability. In: 2011 14th Euromicro Conference on Digital System Design. IEEE. pp. 809–813
24. Pachghare VK (2019) Cryptography and information security. PHI Learning Pvt. Ltd.
25. Sabeel U, Heydari SS, Mohanka H, Bendhaou Y, Elgazzar K, El-Khatib K (2019). Evaluation of deep learning in detecting unknown network attacks. In 2019 International conference on smart applications, communications and networking (SmartNets). IEEE. pp. 1–6
26. Sgaglione L, Coppolino L, D'Antonio S, Mazzeo G, Romano L, Cotroneo D, Scognamiglio A (2019). Privacy preserving intrusion detection via homomorphic encryption. In 2019 IEEE 28th International conference on enabling technologies: infrastructure for collaborative enterprises (WETICE). IEEE. pp. 321–326
27. Sharma N (2017) A Review of Information Security using Cryptography Technique. International Journal of Advanced Research in Computer Science, 8(4).
28. Sowjanya K, Dasgupta M (2020) A ciphertext-policy attribute based encryption scheme for wireless body area networks based on ECC. J Inf Secur Appl 54:102559
29. Stinson DR, Paterson MB (2017) Cryptography: theory and practice. CRC Press, Boca Raton
30. Turan MS, Perlner R, Bassham LE, Burr W, Chang D, Jen Chang S, Dworkin M, Kelsey J, Paul S, Peralta R (2011) Status report on the second round of the SHA-3 cryptographic hash algorithm competition. NIST Interagency Report, 7764.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Anumukonda Naga Seshu Kumar[1] · Rajesh Kumar Yadav[1] · Nallanthighal Srinivasa Raghava[2]

✉ Anumukonda Naga Seshu Kumar
seshu1203@gmail.com

[1] Department of Computer Science and Engineering, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India

[2] Department of Electronics and Communication Engineering, Delhi Technological University (Formerly Delhi College of Engineering), New Delhi, India

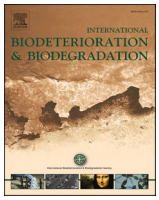
International Biodeterioration & Biodegradation 193 (2024) 105852



Contents lists available at ScienceDirect

International Biodeterioration & Biodegradation

journal homepage: www.elsevier.com/locate/ibiod



Enhancement in sugar extraction from *Pistia stratiotes* through statistical optimization of alkaline pre-treatment and enzymatic hydrolysis

Sweeti Mann [a], Jai Gopal Sharma [a], Rashmi Kataria [b,*]

[a] Department of Biotechnology, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi, India
[b] School of Biosciences and Technology, Vellore Institute of Technology, Vellore, Tamil Nadu, 632014, India

ARTICLE INFO

Keywords:
Alkaline pre-treatment
Cellulose
Enzymatic hydrolysis
Saccharification
Optimization

ABSTRACT

In lignocellulosic biomass, reducing structural recalcitrance and enhancing hydrolysis efficiency are crucial factors for increasing fermentable sugars and the production of valuable products. This biomass substrate comprises lignin, hemicellulose, and cellulose. In this study, response surface methodology was employed to optimise alkaline pre-treatment followed by enzymatic hydrolysis, aiming to enhance the saccharification of *Pistia stratiotes*. The NaOH concentration during pre-treatment significantly influenced the delignification process, resulting in increased cellulose content. The highest cellulose content was achieved with 2.47% NaOH at 120 °C for 60 min, leading to enhanced cell porosity and facilitating greater enzyme saccharification accessibility. Under these optimized conditions, the sample exhibited a 51.66% cellulose content. The physicochemical characteristics of the cellulose obtained after pre-treatment were analysed using SEM, FTIR, and TGA. After enzymatic hydrolysis of the cellulose with a commercially available cellulase enzyme, 31.06 g/L of reduced sugar was produced after 72 h. This study demonstrates that alkaline pre-treatment of *Pistia stratiotes* significantly increased its cellulose content, leading to a higher sugar yield during enzymatic hydrolysis.

1. Introduction

The primary issue facing the ecological niche that negatively affects aquatic and terrestrial life systems is environmental pollution. Fossil fuel consumption is the primary source of this pollution. The majority of environmental contaminants are hydrophobic so that they dissolve slowly in the water, which restrict microorganisms to utilize them as carbon source.(Gu, 2016).

The need for energy has grown significantly in recent years, placing additional strain on the use of fossil fuels, a resource that is running out quickly. In order to meet the demands of the world's expanding population, significant production and resource extraction will continue to accelerate due to the rising consumption of natural resources, particularly the non-renewable ones, which power most countries' economies today (Gu, 2020). Many scientists are employed worldwide in the field of waste-based renewable energy production By using greener and more innovative methods, renewable energy sources are displacing fossil fuels. Biofuel is a form of renewable energy obtained from sources such as agricultural waste, food waste, algae and municipal waste (Wannapokin et al., 2018). One likely way to achieve the Sustainable Development Goals (SDGs) is through the production of biofuels from lignocellulosic biomass (LCBs) (Nazari et al., 2021). LCBs contain secondary cell wall enriched with sugars that are feasible, sustainable and economically directed to the production of valuable products such as biofuels, polymers and biochemicals (Culaba et al., 2022). Currently United States and Brazil are front runners in bioethanol production using corn as a substrate, thereby creating the deficiency of food crops (Tse et al., 2021). Both on land and in water, a variety of invasive species can be used as a substrate for the long-term, sustainable production of valuable products. These lignocellulosic invasive weed species serve as an alternative to food crops in the same capacity. A lot of work has gone into using terrestrial lignocellulosic substrates to produce reduced sugars that microbes can use to produce biofuels and biogas (Moerman, 1996). The inability of lignocellulosic biomass to grow quickly due to a lack of agricultural resources and land presents a barrier to the industrial production of bioenergy (Chen et al., 2015). Aquatic biomass could be utilised as a substrate for the synthesis of biofuel to address the aforementioned issue. The advantage of aquatic weed over terrestrial lignocellulosic biomass is its lower lignin content and faster growth rate. Because of their components, a variety of aquatic plants have the

* Corresponding author.
E-mail addresses: sweeti.dhankhar.1988@gmail.com (S. Mann), sharmajaigopal@gmail.com (J.G. Sharma), rashmikataria@gmail.com, rashmi.kataria@vit.ac.in (R. Kataria).

https://doi.org/10.1016/j.ibiod.2024.105852
Received 20 January 2024; Received in revised form 17 June 2024; Accepted 2 July 2024
Available online 8 July 2024
0964-8305/© 2024 Published by Elsevier Ltd.

potential to produce bioethanol.

Pistia stratiotes (water lettuce) is a monocotyledonous weed found in fresh water. It can obstruct the water channel due to its dense growth connected with the root nodules. This weed's primary drawback is that it can deteriorate the quality of the water, which is why it's considered an invasive weed. When used as a substrate for the production of bioethanol, this weed's drawback can be turned into a useful resource for financial gain (Jayanth, 2000). Early in the 1990s, *Pistia stratiotes* was discovered to be a valuable carbon source for the creation of products like biogas. In addition to the weed's anaerobic digestion, which yields biogas, continuous fermentation also forms an alternative by producing bioethanol. (Yang et al., 2022). Harvesting the aquatic weeds is the first stage in turning them into biofuels. Compared to alternative strategies, mechanically more efficient methods can be used to remove these weeds (Alam et al., 2021).

Lignocellulosic substrates are highly recalcitrant as a consequence of lignin along with cellulose and hemicellulose in the cell wall of plants. For the saccharifying enzymes to access the substrate, this is a significant barrier. Pre-treatment is necessary to get rid of lignin and hemicellulose and make cellulose more accessible to hydrolysing enzymes in order to improve the saccharification from lignocellulosic biomass (Fillat et al., 2017). Pre-treatment raises the biomass porosity by eliminating lignin and hemicellulose (Asghar et al., 2015). Pre-treatment can be categorized as biological (fungi) like white-rot fungus (Saha et al., 2016), physical (milling, extrusion, microwave heating), chemical (alkali, acidic, ionic, and with organic solvents) and physiological (CO_2 explosion and wet oxidation)(Aguilar-Reynosa et al., 2017). However, none of these meet the requirements in terms of inherent advantages and drawbacks. Thus, to improve the effectiveness of the conversion process, a valuable pre-treatment procedure is evaluative. When fungi are co-cultivated during a biological pre-treatment, their combined growth can sometimes inhibit the growth of the individual fungi, but it also has a synergistic effect on lignin degradation and the expression of lignolytic enzymes (Meehnian et al., 2017).

Alkaline pre-treatment is a more intently studied technique and has many advantages, such as effective lignin solubilization in contrast to other pre-treatments. Alkaline pre-treatment is used for delignification with minimal loss of reduced sugar and unescorted by inhibitory compound formation (Alvira et al., 2010). Saphonication and solvation are the causes of swelling in alkaline pre-treatment. The lignin and hemicellulose eater bonds break down as a result of the saponification reaction. The surface area of cellulose increases as the polymerization and crystallinity decrease as a result of the swelling of the biomass. Alkaline concentration, reaction temperature, and reaction time are some of the parameters that need to be optimized in order to get the most out of the alkaline pre-treatment (Kim and Han, 2012).

Response surface methodology (RSM), a statistical tool, is used for the optimization process, which involves combining the results of each experiment to optimise multiple variables under various experimental conditions (Ali et al., 2015; Manmai et al., 2020). It is a statistical tool that draws on scientific techniques such as analytical and classical methods, producing responses that are influenced by a variety of variables and that ultimately relate to optimization. Numerous lignocellulosic biomass, including rice straw grass, switchgrass, and corncob, are optimized using RSM (Başar et al., 2021; Gundupalli et al., 2022; Selvakumar et al., 2022). The most popular and frequently utilised RSM design is the central composite design (CCD). Enzymatic hydrolysis requires cellulase enzyme for conversion of cellulose into reduced sugar. The source of carbon has the lowest concentration limit at which the balance between newly produced and dead cells keeps the amount of living cells relatively constant (Gao and Gu, 2021).

Therefore, the main goal of this investigation was to optimise the alkaline pre-treatment of *P. stratiotes* starting from raw material to enzymatic hydrolysis. To our knowledge, no pretreatment study on *P. stratiotes* has been reported. The content of cellulose is affected by three distinct factors: temperature, time, and alkali concentration. Central composite design (CCD) and factorial design were used to optimise the pre-treatment process. Fourier transform infrared spectra (FTIR), thermogravimetric analysis (TGA), and scanning electron microscopy (SEM) were used to further characterise the recovered cellulose fraction following the pre-treatment. The commercially available cellulase enzyme was used to hydrolyse the recovered solid fraction (cellulose). The estimation of reduced sugar was computed after hydrolysis. It is possible to conclude from the study's observations that this is the first study to optimise the alkaline pre-treatment of Pistia stratiotes using RSM, resulting in a higher cellulosic content.

2. Materials and methods

2.1. Collection and preparation of sample

Pistia stratiotes, commonly known as water lettuce, were harvested in August 2022 from a local pond in Delhi, India's Hauz Rani City Forest (28.51590 N, 77.21110 E). The harvested material included all plant parts: roots, stems, and leaves, with an average plant size of 35 cm. Subsequent to collection, the sample underwent a dual cleansing process with tap water to eliminate any impurities. Following this, it was allowed to air-dry for 48 h at room temperature. The air-dried sample was further dehydrated in an oven at 45 °C overnight and subsequently ground using a grinder. The resulting finely ground sample was stored in an airtight polybag for future use.

2.2. Sample composition analysis

2.2.1. Quantification of cellulose

To estimate cellulose content, the Updegraff method was employed. Initially, for washing 70 mg sample was incubated in 1.5 ml of 70 % ethanol at 70 °C for 1 h and then removed out ethanol by an aspirator twice. After that 1 ml of acetone was added for 2 min and then aspirated out and put the sample in a fume hood for 3–4 h, keep the sample in the oven at 37 °C overnight, resulting in what is known as alcohol-insoluble residue (AIR). This sample was stored at room temperature for further use. This AIR was then subjected to a 30-min treatment at 100 °C in a water bath using 3 ml Acetic nitric reagent (80% glacial acetic acid: 10 % nitric acid: 20 % Milli-Q). Following this treatment, the sample underwent rinsing with ethanol and water until cellulose became apparent. After that 4 ml acetone was added to the sample and evaporated in a fume hood. The resulting residue was subsequently placed in an oven overnight at 37 °C. Subsequently, 1 ml of 67 % H_2SO_4 was added to the sample, and the mixture was placed in a shaker for thorough mixing. After the cellulose was effectively mixed with the acidic solution then glucose standards were made from 100 mg/ml stock solution. Anthrone reagent was prepared by mixing 0.3 % anthrone in H_2SO_4. For testing 500 μl sample and blank take out in the Eppendorf and add 1 ml anthrone reagent in each, now boil each for 5 min and take the OD at 620 nm. The anthrone test was employed to estimate the amount of glucose present in the cellulose (Kumar and Turner, 2015).

2.2.2. Hemicellulose and lignin

The fibre analysis method was used to estimate hemicellulose. This methodology starts with the estimation of neutral detergent fibre (NDF), moves on to the calculation of acid detergent fibre (ADF), and subtracts ADF from NDF to find the concentration of hemicellulose (Holtzapple, 2003). National Renewable Energy Laboratory (NREL) protocols are followed for the lignin estimation. The first step in this process was the extraction of the sample with the help of milli Q water using probe sonicator (Branson Sonifier W450 Digital, tip size 6.5 mm) (Zhou et al., 2017). The resulting extracted sample was treated with 72 % H_2SO_4 in the water bath for 1 h. It was then diluted with 4 % milli Q and autoclaved for 1 h at 121 °C and 15 psi. After that, the sample was filtered using a vacuum filtration unit in a crucible that had been preheated. After being dried at 105 °C for 4 h, the residue was weighed and

designated as acid-insoluble residue (AIR). Then the crucible was transferred into muffle furnace at 575 °C for 24 h and weighed once more. The outcome obtained after burning the sample in the muffle furnace gives % of ash in filter (Xia et al., 2022). Acid-soluble lignin was calculated by measuring absorbance at 280 nm (Nomanbhay et al., 2013). All the calculations were performed using the formulas mentioned below Eqs. (A.1, A.2 and A.3).

Eq.(A) was used to calculate the lignin.

$$\%\ AIR = \frac{(Crucible\ plus\ AIR\ weight - Crucible\ weight)}{Sample\ loading\ in\ Crucible} \times 100$$

$$\%\ Ash\ in\ filter = \frac{(Crucible\ plus\ ash\ weight - Crucible\ weight)}{Sample\ weight} \times 100$$

$$\%\ Acid\ insoluble\ lignin = \%\ AIR - \%\ Ash\ in\ filter \quad \text{Eq.(A.1)}$$

$$\%\ Acid\ soluble\ lignin = \frac{UV_{absorbance\ (280nm)} \times Voulme\ of\ filtrate \times dilution}{Absorptivity \times dry\ weight \times pathlength} \quad \text{Eq. (A.2)}$$

$$\%\ Lignin = \%\ Acid\ insoluble\ lignin + \%\ Acid\ soluble\ lignin \quad \text{Eq. (A.3)}$$

2.3. Sample characterization

2.3.1. Fourier-transform infrared (FTIR) spectroscopy

The changes in functional groups after alkaline pre-treatment were determined through the use of FTIR spectroscopy (PerkinElmer 400 FTIR/FTIR). With the control sample, the spectra span the frequency range of 4000 cm^{-1} to 400 cm^{-1}. Pellet preparation involved the use of potassium bromide (KBr), and scanning was done at a resolution of 4 cm^{-1} (Kataria et al., 2018a).

2.3.2. Thermogravimetric analysis (TGA)

The TGA of raw and alkaline pre-treated samples was performed by PerkinElmer TGA 4000. The samples were heated using a nitrogen flow rate of 10 mL/min and a heating range of 10 °C min^{-1} from room temperature to 600 °C (Kataria et al., 2018b; Umesh et al., 2022)

2.3.3. Scanning electron microscopy (SEM)

The changes in surface and fibre morphology of pre-treated sample were examined using Scanning Electron Microscopy (SEM) (Model: EV018 Zeiss, Germany) with 5 kV voltage at 10,000× magnification. SEM analysis was used to identify the structural alterations following cell wall breakdown (Umesh et al., 2022).

2.4. Optimization of pre-treatment using RSM

RSM is a multivariant statistical tool that provides a new approach for determining the ideal pre-treatment state. Design Expert software assisted in the development of a central composite design, which was used to determine the effects of various independent variables on cellulose extraction. Temperature (60, 90, and 120 °C), time (30, 45, and 60 min), and NaOH concentration (0.5, 1.75, and 3 %), were the variables. RSM was responsible for determining the results. The variables in this model were represented by codes, which were represented by the numbers −1,0, and 1. The neighbouring distance from the value of the central point determines how variables are coded. The experiment's independent variables, such as the temperature, time, and NaOH concentration, were transformed into code variables Xi (Eq. (B.1))

$$Xi = 2\left(\frac{OV - V}{\Delta}\right) \quad \text{Eq. (B.1)}$$

Equation 2 comprises the following terms:

Δ (difference between largest and smallest values (range)), OV (original variable), and V (average of the largest and smallest values of variables or mid value). Three independent variables with lower and higher levels are mentioned in Table (A.1).

2.4.1. Statistical analysis and model fitting

The statistical tool known as Analysis of variance (ANOVA) was employed for divination of the statistical factors required in the assessment of the RSM model between the independent variables and response of parameters. Utilising the software Design Expert, regression analysis of trial data was employed to generate the contour plot, pareto chart, and optimized condition. The influence of independent variables on the cellulose concentration is shown by the contour plot and pareto chart. The outcome of the model predicted statistically significant values by F test and p value. If the value was less than 0.05, then the model was stated as statistically significant with lack of fit test that is insignificant in the same model. Independent factors significantly impacting the response were determined by p test value less than 0.05 with confidence level above 95 %. Significance of the model was accessed by R^2 predicted and R^2 adjusted value (Ramaraj and Unpaprom, 2019). To determine the ideal condition divided into linear, quadratic, and interactive components, the effects of independent variables on the dependent variable (response) were estimated using a polynomial equation. Eq. (C.1) displayed the entire cellulose concentration estimation.

$$Y = \beta_o + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \sum_{i<j}^{k} \beta_{ij} X_i X_j + \epsilon \quad \text{Eq. (C.1)}$$

where Y is the predicted response (cellulose concentration); β_o the intercept; β_i the linear constant; β_{ii} the quadratic coefficients; β_{ij} the interaction effect of coefficients; X_i and X_j the coded values used for variable parameters; and ϵ is the random error (Dahunsi et al., 2019).

2.5. Enzymatic hydrolysis

After pre-treatment, 10 ml of citrate buffer (0.05M, pH 4.8) was added to 150 ml test tubes containing 0.5 g of both pre- and untreated samples for enzymatic hydrolysis. The commercially available cellulase enzyme (40000 U/g) was sourced from IndiaMart, and the Filter Paper Unit (FPU) was calculated following NREL protocols. The cellulose loading was adjusted to 25, 50, and 75 FPU/g of dry biomass.

The hydrolysis was carried out in an incubator shaker (New Brunswick Innova 44 series, Germany) with a rotation speed of 150 rpm for 96 h at 50 °C. Samples were withdrawn at 24, 46, 72, and 96 h. The reducing sugar was quantified using the DNSA method as described by Liu et al. (2021). After enzymatic hydrolysis, the sugar yield was determined using Eq. (D.1), and each experiment was conducted in triplicate.

$$Sugar\ yield\ (\%) = \frac{0.9\ X\ reducing\ sugar\ \left(\frac{g}{L}\right) X\ volume\ (L)}{solid\ cellulose\ fibre\ wt.\ (g)}\ X\ 100 \quad \text{Eq. (D.1)}$$

2.6. Reducing sugar extraction

The 3, 5-Dinitrosalicylic acid (DNSA) method was used to estimate the total amount of reducing sugar (Miller, 1959). After preparation of DNSA reagent (10 g 3,5 DNS, 0.5 g sodium sulphate, 182 g sodium potassium tartrate, 10 g sodium hydroxide and 2 g phenol in 1 L), it was kept in an amber bottle for storage. For the standard curve, a stock solution of glucose standard was prepared. Initially, 1 ml DNSA reagent was added to the sample, glucose standards and a blank and put in the water bath and left for 15 min. Then the samples were removed from the water bath and 5 ml milli Q was added for dilution to take the OD of each sample to quantify reduced sugars, the optical density of the samples was measured at 540 nm using a glucose standard.

3. Results

3.1. Composition analysis of the sample

P. stratiotes can be used as a good carbon source for the production of value-added products. In the raw sample, the estimated contents of cellulose, hemicellulose, and lignin were 25.90 %, 18.44 % and 25.25 % respectively. The composition analysis proves that in this weed, a high amount of cellulose content was found compared to hemicellulose and lignin. After the chemical pre-treatment, the composition analysis showed that it might be a useful carbon source for the production of bioethanol. By applying the alkali pre-treatment to remove the lignin content from the raw sample, the cellulose content could be increased.

3.2. Optimization of alkali pre-treatment and validation

After pre-treating *P. stratiotes* with alkali (NaOH), which had an impact on the raw material's depolymerization, cellulose was the final product, with lignin removed. The reduced sugar obtained from the hydrolysis of cellulose can be used to further produce ethanol. The best conditions were achieved by optimising biomass using RSM, which is 2.47 (w/v) NaOH at 120 °C for 60 min after autoclaving. The optimal condition gives the result of 51.66 % cellulose. The result of the alkali pre-treatment under various conditions is cellulose, which is shown in Table (A.2).

The factors that affect the expected changes in the mean response when the factors change from a lower to a higher level, which also establishes the coded coefficients. The alkali concentration shows major positive impact so that this parameter has an effect on increased response (cellulose concentration) proceeding from 0.5 % to 3 %. If the alkali concentration is higher, then the cellulose concentration will also be enhanced. However other factors like time, temperature, alkali conc. *alkali conc. And alkali conc.*time, shows the negative sign which successively impacts on reducing the response.

Table (B) illustrates an ANOVA statistical model for cellulose estimation from CCD. ANOVA and the lack of fit test are used to analyse the model's fitness. When the experimental data fits the model, significant regression and non-significant lack of fit are displayed. This model's statistical advantage was evaluated and attributed to the interactions between the model's factors and p-value. Table (B) shows that the model's p-value for cellulose concentration is 0.000, which is less than 0.050 and suggests that the response to the model is substantial and meaningful. The model is considered highly significant when the p-value is less than 0.001. A p-value greater than 0.1 indicates that the model is not significant. However, because of the large f value (53.79), the model result has a significant impact on the response. This f-value is probably going to demonstrate the significance of the model. There is only a 0.01 % chance in the statistical model that a large f value could arise from (Pashaei et al., 2020).

3.2.1. Model summary

The adjusted R^2 of 96.15 % and the predicted R^2 value of 92.48 % agree rationally, indicating a difference of less than 4 %. This result demonstrates that this model can accurately represent the percentage of cellulose extraction from *P. stratiotes* and can be shown as coded variables, as shown in Eq. (E.1).

3.2.2. Regression equation in Uncoded units

Cellulose = 64.90 + 12.18 Alkali conc. - 0.986 Time - 0.697 Temp.

−2.409 Alkali conc.*Alkali conc.+ 0.00465 Time × Time +0.001267 Temp.*Temp.

−0.1395 Alkali conc.*Time +0.0675 Alkali conc.*Temp. +0.01109 Time × Temp. Eq.(E.1)

3.2.3. Contour plot

The contour plot describes the impact of different variables on the response, and change in cellulose content depending on the temperature and time). The effects of temperature and time on cellulose are interpreted in Fig. (A.1). A notable increase in cellulose is seen as a result of the temperature rising over time. A rise in cellulose content of more than 50% was noted after 55–60 min at 110–120 °C. Conversely, after pre-treatment, a higher amount of cellulose content is obtained by raising the temperature and alkali concentration. The maximum cellulose content, or >50%, is achieved at 120 °C and >2.5 % alkali content, as Fig. (A.2) illustrates.

3.2.4. Pareto chart

The Pareto chart was used to clearly illustrate the significance and magnitude of the impact of independent variables on response. The complete values of the regulated effects, arranged from highest effect to lowest effect, are shown in the Pareto chart. The chart's reference line indicates statistical significance for the effect. The statistical significance is shown by the bar that crosses the reference line. The cellulose Pareto chart is shown in Fig. (A.3). This chart indicates that the parameters are statistically significant because of time × temperature (BC), alkali conc.* time (AB), alkali conc.* temp. (AC), alkali conc. (A), alkali conc.* alkali conc. (AA), temp. (C), and time (B) cross the reference line.

3.2.5. Response (cellulose) optimization

The response observed in the complete process was cellulose. After statistical analysis of the optimization model, the target cellulose content was 50.87 %, and the objective was to maximise the cellulose outcome with a lower cellulose content of 21.17 %. The ideal condition was discovered at 2.4 % alkali concentration with 60 min at 120 °C temperature after multiple responses. Post the optimization experiment the cellulose fit was 51.65 % with 1 % composite desirability.

3.2.6. Response outcome

RSM was used to determine the response (cellulose content) outcome, and it fit 51.66 % of the cellulose content with a standard error of 1.06. With a prediction interval ranging from 47.92 to 55.39 %, the confidence interval fell between the range of 49.92–54.01 % cellulose content.

3.2.7. Multiple response prediction

The optimized condition was discovered through multiple response prediction after carrying out a series of experiments. All variables were then fixed, and after 60 min at 120 °C, the maximum cellulose content was discovered at 2.46 % alkali content.

3.2.8. Optimization plot

Optimization of alkaline pre-treatment was performed, optimized condition considered at 2.47 % NaOH, 120 °C with 60 min. This optimized condition was performed separately to check the reliability. After performing the optimized condition, 51.67 % cellulose was found in the hydrolysate which shows the reliability of the model. A model created by Minitab software in which independent variables are adjoined to estimate the target response. Combined desirability ranges from 0 to 1 are helpful for the calculation of optimization. Cellulose has an independent desirability outcome of 1.0 as a predicted outcome of 51.65 %. The separate desirability of 1.0 is an excellent score which shows that response is immediate to their absolute setting. The optimization of response (cellulose) gives an ideal desirability score as cellulose was an absolute setting that is included in the justifiable range. The optimized condition established after the RSM model is represented in Fig. (A.4).

3.3. Effect of cellulase loading on enzymatic hydrolysis

A comparison of the pre-treated and raw samples' enzymatic hydrolysis shows promising results following pre-treatment. 25 FPU, 50 FPU, and 75 FPU cellulase loadings are used to hydrolyse the raw and pre-treated samples in a shaker at 150 rpm and 50 °C. The sample that was pre-treated and hydrolyzed by 50 FPU of cellulase, displayed a maximum concentration of 25 g/L of sugar reduction (fig. (B.2). With an enzymatic activity of 75 FPU, the pre-treated sample's maximum reduced sugar of 33.62 g/L was formed in 72 h. After 72 h, the pre-treated sample's maximum sugar yield was discovered, with 75 FPU enzymatic activity of 60.53% (Fig. (B.2). In contrast, under the same conditions, the raw sample yields 6.68 g/L of reducing sugar with 50 FPU in 72 h following hydrolysis. With 75 FPU, the maximum reduced sugar of 11.56 g/L was formed in 96 h. After 96 h, the raw sample's maximum sugar yield of 20.80 % was formed, with an enzymatic activity of 75 FPU Fig. (B.1). It demonstrates how successful the pre-treatment is when hydrolysis occurs. Mass balance of the complete study was determined by the total product formation from the substrate (Njoku et al., 2013). The mass balance is shown in Fig. (C).

3.4. Improvisation of the raw and treated sample

3.4.1. Phonological modulation in raw material (SEM analysis)

The sample's initially smooth surface transforms into a rough texture under the influence of elevated temperatures and alkaline conditions, causing the breakdown of cellulose and hemicellulose bonds upon lignin removal. This alteration increases the surface area of the biomass, enhancing its accessibility to enzymes, as demonstrated by Kataria et al. (2017). Post-alkaline treatment, noticeable structural changes in the untreated sample are evident, as depicted in Fig. (D). Comparative analysis with the untreated raw material reveals a more pronounced disruption in the sample following alkaline pre-treatment. In Figure (D.2), the most significant disruption is observed after alkaline pre-treatment. Notably, severe conditions (3 % NaOH concentration) also exhibit substantial disruptions, as illustrated in Fig. (D.4) In contrast, mild pre-treatment conditions (0.5 % NaOH concentration) are shown in Fig. (D.3).

3.4.2. FTIR spectrum of the untreated and treated sample

The functional group and its properties are specifically revealed by the FTIR spectrum. The treated sample's FTIR spectra (Fig. E.1) showed a broad, highly intense peak spanning 3500–3200 cm^{-1} that indicated N–H stretching of primary amines, an increase in free O–H bonds as a result of extractives being removed, and the presence of cellulose (Lu and Hsieh, 2010). When the sample is left untreated, an ester bond-related band at 1740 cm^{-1} indicates that the polymer within is mutually interconnected; however, when the sample is treated with alkali, xylose solubilization takes place and the band vanishes (Sills and Gossett, 2012). The treated sample exhibits more intense peaks at 1430 cm^{-1} and 1320 cm^{-1}, which correspond to C–H_2 and C–H bending, respectively, in relation to cellulose (Sombatpraiwan et al., 2019). The hemicellulose acetyl ester's stretching vibration is reduced in the treated sample, indicating the removal of the xyloglucan acetyl group with the COOH group from the lignin hemicellulose matrix. This is represented by the peak at 1231 cm^{-1} (Trevorah and Othman, 2015) 2015. The peak at 1024 cm^{-1} represents polysaccharides. The treated sample has high intensity at this region because of C–O, C═C, and C–C–O stretching for polysaccharides, which correlate with hemicellulose decomposition (Bano and Irfan, 2019). The aromatic band ranges from 950 to 700 cm^{-1}, which corresponds to the β-glycosidic linkage in cellulose and hemicellulose units (Gusain and Suthar, 2017). Due to C–O–C stretching in the β-glycosidic linkage in cellulose and hemicellulose, which makes cellulose accessible to enzymes, the intensity of the peak at 880 cm^{-1} increases in the treated sample (Deng et al., 2019).

3.4.3. Thermogravimetric analysis (TGA) and differential thermogravimetric analysis (DTG)

Fig. (E.2) shows the thermogravimetric analysis of the untreated sample (raw), mild (0.5 % NaOH conc., 30 min and 60 °C), harsh (3 % NaOH Conc., 60 min and 120 °C), and optimized pre-treated condition (2.47 % NaOH conc., 60 min and 120 °C) to determine their degradation characteristics at 0–600 °C. During the TGA analysis shown in Fig. E.2, four distinct types of weight reduction were observed in *Pistia stratiotes*: moisture, cellulose, hemicellulose, and lignin reduction. Water evaporation causes weight loss to begin at 100 °C. At 200–300 °C, hemicellulose depolymerizes and breaks down its cellulosic glycosidic bonds. Water loss causes a mild, raw sample to lose 10–12 % of its weight between 100 and 200 °C. Unexpected weight loss of approximately 50 % was noted in all samples between 250 and 350 °C as a result of hemicellulose, cellulose, and lignin degradation. Lignin breaks down between 200 and 500 °C. The primary thermal degradation region was observed to be between ~325 and 345 °C, indicating the degradation of cellulose and lignin in the sample. (Basak et al., 1993). The range of 200–500 °C was where the optimized condition's 55 % weight loss occurred. This indicates the sample's delignification and is the result of the alkaline pre-treatment. The optimized sample's DTG curves show a peak at 345 °C, respectively, with the hydrolyzed sample showing the strongest signal because of its higher cellulose content. The raw material's maximum thermal stability, which corresponds to its non-cellulosic content, was measured at 325 °C. Fig. (E.3) illustrates the optimized sample's high thermal stability, with the DTG curve indicating a maximum stability temperature of 345 °C. The optimized sample's degradation temperature is higher than that of other lignocellulosic materials that have been reported, such as Napier grass (300 °C) (Reddy et al., 2018), roselle fibres (326 °C)(Kian et al., 2017) and kans grass (340 °C) (Baruah et al., 2020).

4. Discussion

Biofuels can be produced from biomass, which is an environmentally friendly way to mitigate the energy crisis. The sustainable development goal is aided by the bioconversion of lignocellulosic biomass into biogas and biofuel, which lessens the carbon footprints. Lignocellulosic biomasses are composed of lignin, hemicellulose, and crystalline cellulose. Unlike the first generation, which used sugar directly, intrinsic biomass recalcitrance significantly hinders lignocellulose's bioconversion by protecting carbohydrates from degradation. The recalcitrance of wrapped cellulose significantly hinders the recovery of fermentable sugar from it. These types of biomass degrade slowly, so effective pre-treatment and enzyme dosage are necessary (Hu et al., 2023a). Pre-treatment is a crucial stage in the bioprocess technology process as a whole. In modern biorefineries, various pre-treatments such as physical, chemical, and biological are used to overcome the recalcitration of lignocellulosic biomass (Mankar et al., 2021). Physical pretreatment is used to reduce the size of particles and increase the surface area of lignocellulosic biomass by disrupting of structure of cellulose, hemicellulose and lignin for accessibility of enzymes e.g. grinding (Ji et al., 2017) Chemical pretreatment is more effective for increasing the hydrolysis and it is a simple procedure with high efficiency. Different chemicals are used in this process like acids, alkali, organic solvents and ionic liquids, all of these can break down the lignocellulosic biomass. Acidic pretreatment is specific for lignin degradation by using different concentrations of acid, reaction time and temperature. The disadvantage of this pretreatment is its ineffectiveness in the lignin removal and stopping its hindrance effect on cellulose (Rodrigues Gurgel da Silva et al., 2019). Alkali pretreatment is helpful for lignin and hemicellulose removal and increases the accessibility to the enzymes for hydrolysis with a high reduced sugar yield. The expensive neutralization and secondary contamination are its drawbacks (Malik et al., 2020) Ionic liquids are advantageous over acids with non-flammability, thermal stability and chemical adjustability (Abushammala and Mao, 2020).

Biological methods use fungi and bacteria, its advantage over other methods is the low cost of the downstream process, simple procedure, and low energy consumption. The drawback is its low-efficiency limit and time-consuming process (Rezania et al., 2020)(Li et al., 2022). Lignocellulosic biomass includes corn stover, which undergoes enzymatic hydrolysis and pre-treatment before being transformed into a value-added product (Li et al., 2017). One abundant and renewable feedstock for making biofuels is grass. Its growth cycle is short, with low lignin content, and its crystallinity is also less. However, getting a high yield of biofuels from grass is difficult due to the stiff cellulose-hemicellulose-lignin network that resists enzyme attack. Consequently, various pre-treatment methods have been applied to different types of grass in order to boost the yield of biofuels (Jin et al., 2015). Due to their low lignin content, aquatic weeds are classified in the same category as grasses. Numerous aquatic weeds, such as *Eichhornia* sp., *Pistia stratiotes*, *Lemna minor*, and *Lemna gibba*, can serve as good carbon sources for the synthesis of biofuel (Gusain and Suthar, 2017). Each of these weeds has a different composition. According to a previous study, duckweed (*Lemna minor*) contains cellulose 55.2%, hemicellulose 32.6 %, and lignin 12.2% (Yadav et al., 2017). The composition of water hyacinth before pre-treatment showed 1.77% acid-soluble lignin, 6.33% acid-insoluble lignin, 32.84% cellulose, and 24.7% hemicellulose. The results of the thermal pre-treatment were different: by using the autoclave technique, 1.58% acid-soluble lignin, 8.35 % acid-insoluble lignin, 27.8% hemicellulose, 29.26% cellulose were formed (Barua and Kalamdhad, 2017). Organic waste also includes rice straw that is used to optimise the alkali pre-treatment process. The temperature, time, NaOH concentration, and solid loading were the independent variables. The sugar reached its peak at 121 °C, 40 min, 2% w/v NaOH concentration, and 5% loading. With 50.5 g/L of glucose, 13.5 g/L of xylose, and 1.3 g/L of arabinose, the maximum sugar content was 65.3 g/L(Valles et al., 2021). A comparative investigation of diverse chemical pretreatment was performed on sorghum straw. Six different pretreatment were applied to it, namely, 2 % (w/v) NaOH, 2 % H_2SO_4, 2 % (w/v) Na_2CO_3, 2 % (w/v) oxalic acid, 2.15 % (w/v) H_2O_2 and 95% (v/v) glycerol pretreatment. In this study, 2 % NaOH pretreatment was most effective for lignin removal and enzyme-mediated hydrolysis process (Bhati and Sharma, 2023). Another optimization of alkaline pretreatment by central composite design with enzymatic hydrolysis study was performed on the cocoa pod husk. The optimal condition was 5%(w/v) NaOH for 30 min at 120 °C increased the content of cellulose from 27.68 (untreated) to 57 % and then enzymatic hydrolysis was performed resulting in 66.80 g/L reducing sugar yield up to 98.75 % (Hernández-Mendoza et al., 2021) which is high from our study. When alkaline pretreatment was done on the *Brassica juncea* with a low concentration of NaOH (0.5 M) at 160 °C for 30 min with 10 % loading. The delignification occurred in the sample and then enzymatic hydrolysis was done by the mixture of cellulase obtained from *A. niger* MTCC284, *T. harzianum*MTCC8230 and *F. incarnatum*KU377454. Then 62.35 mg/ml reduced sugar was formed after hydrolysis in 48 h at 50 °C with 15 % loading (Pant et al., 2021). Optimization of alkali pretreatment was performed with the help of response surface methodology and maximum cellulose content was formed by 2 % NaOH i.e. 62.7 % and highest 78.3 g/L glucose was obtained at 20 g/L enzyme loading with 50 °C for 72 h (Punia and Singh, 2024). Another study on *C. oleifera* fruit shell and its seed cake used as a substrate for enzymatic hydrolysis after oxalic acid pretreatment. SSF (Solid state fermentation) performance of *A. niger* with 50 °C at 200 rpm for 24 h gives 20.58 g/L reduced sugar which is less than our study (Dessie et al., 2024). *Pistia stratiotes* contains 49.4 % carbohydrates, 17.8 % fibres, and 16.5 % protein with 23.8 % ash content (Pantawong R., Chuanchai A., Thipbunrat P., Unpaprom Y., 2015). The cellulose content of *Pistia stratiotes* was extracted in a previous study under various conditions using bleaching agents. Four distinct conditions were used to optimise the pre-treatment: (1) 2% sodium chlorite + 2% glacial acetic acid; (2) 4% sodium hypochlorite; (3) 4% hydrogen peroxide + 4% sodium hydroxide; and (4) 4% soapnut solution. The results of all the conditions in the form of cellulose yield as a percentage were 38.92, 25.70, 10.7 and none in the fourth condition (Umesh et al., 2022). In another study, 1% alkali (NaOH) and 1% H_2O_2 were used to treat *Pistia stratiotes* and water hyacinth. Following this treatment, the cellulose results in water hyacinth increased from 19.7 to 34.2% and in *Pistia* from 16.5 to 28.4% (Mishima et al., 2008). After the Pre-treatment, besides ethanol, other solid-liquid products are also produced like foliar fertilisers, jet fuels and catalysts from the *Pistia stratiotes* (Yang et al., 2022). The optimization of pre-treatment was done on water hyacinth. Different acids like HCL/H_2SO_4/HCOOH (2% v/v) with varied concentrations were used such as 1%,2%,3% and 4% v/v to obtain a high amount of reduced sugar. Also, 3% v/v NaOH was used for alkali pre-treatment. The observed results gave maximum reduced sugar at 37.89 mg/100 ml filtrate with 4 % H_2SO_4 (dilute acid pre-treatment). With the alkali pre-treatment, 3% NaOH gave 17.185 mg/100 ml sugar concentration (Awasthi et al., 2013). In contrast, the highest amount of reduced sugar detected in our study was 33.62 g/L following 72 h of enzymatic hydrolysis. The pre-treatment of lignocellulosic biomass with enzymatic hydrolysis is crucial. The saccharification of cellulosic biomass or the low degradation of lignocellulosic biomass for the production of valuable products is the bottleneck in this field, necessitating the use of economically used enzymes that are effective in the saccharification process. Enzymes ought to be less expensive to produce at the industrial scale. The cost of using sugars from the lignocellulosic biomass reaches 25–30% of the total cost of producing biofuel (Hu et al., 2023b). Either commercially available or in-house-produced cellulase enzymes can be used for enzymatic hydrolysis. The primary source of fermentable sugars is plant cell walls. Certain bacteria possess the ability to break down cellulose, hemicellulose, and lignin. For example, the bacterium, *Trabulsiella* sp. is capable of efficiently breaking down the lignin model compound, guaiacylglycerol-β-guaiacyl ether (Suman et al., 2016). *Clostridium thermocellum* is an isolated bacteria which is capable for degradation of the cellulose and used for saccharification process. It is isolated by the goat rumen and is a thermophilic bacteria that is active at temperature of 50–70 °C and works on the agro-industrial waste (Hamann et al., 2015). Apart from bacteria, the fungus is also used for the production of enzymes. Different types of fungus used for the production of cellulase enzymes like *Aspergillus* sp. A1C2-06, *Talaromyces verruculosus* A1C2-05 (Fontes et al., 2023), *Aspergillus niger* (Sulyman et al., 2020) and *Trichoderma resei* (Wu et al., 2019). In the previous study, the *A.niger* produced enzymes by using the waste of *C. oleifera* as a substrate and the conditions were 50 °C at 200 rpm for 24 h and produced 20.58 g/L sugar (Dessie et al., 2024). Another study shows that using the same fungus stain and conditions were different than was at 50 °C at 150 rpm for 48 h on the wheat straw after thermal pretreatment produced 32.90 g/L sugar (Infanzón-Rodríguez et al., 2022) Commercially available cellulase enzymes suitable for the saccharification process are also reasonably priced. We employed the inexpensive, readily available commercial cellulase enzyme in our investigation to achieve successful saccharification results.

5. Conclusions

Alkali pretreatment of biomass is a crucial step for lignocellulosic biomass delignification. This lignin removal process results in the exposure of cellulose and the liberation of sugars. Depending on the type of plant biomass, pretreatment process need to be employed. In the present study, alkaline pretreatment was optimized for *Pistia stratiotes* biomass and 51.66 % cellulose was formed. The biomass characterization of pretreated biomass also revealed the delignification. The enzymatic hydrolysis was done after pretreatment by a commercially available enzyme which gives a maximum of 31.06 g/L reduced sugar with 75 FPU in 72 h. Hence, this renewable sugar obtained by hydrolysis may be employed for various microbial metabolites production such as ethanol, bioplastics, organic acids and different metabolites. This makes it a sustainable and cost-effective process.

Statements and declarations

All the authors declare that the work done in the present manuscript is original and done at Delhi Technological University, Delhi, India.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Sweeti Mann: Writing – original draft, Methodology, Investigation, Formal analysis. **Jai Gopal Sharma:** Writing – review & editing, Supervision, Resources. **Rashmi Kataria:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in the paper. All the authors declare that the work done in the present manuscript is original.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank the senior management of Delhi Technological University (DTU) for their constant support and financial assistance. We are also thankful to All India Institute of Medical Sciences (AIIMS) for SEM Analysis and University Grants Commission (UGC) for providing fellowship support.

Appendix

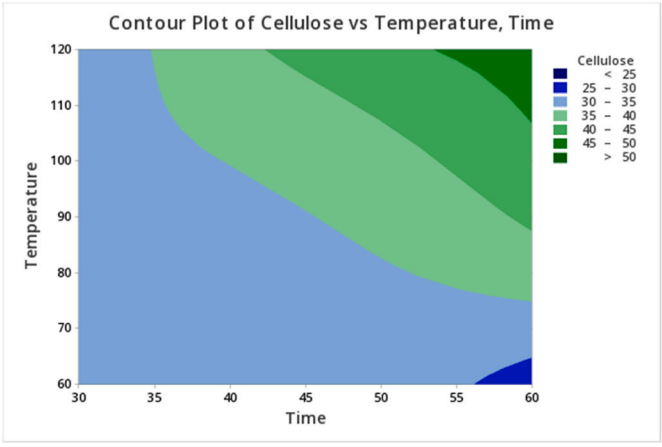


Fig. (A.1). Contour plot shows the effect of temp. And time on the cellulose content. An increase in cellulose content was noticed when temperature increased with time.

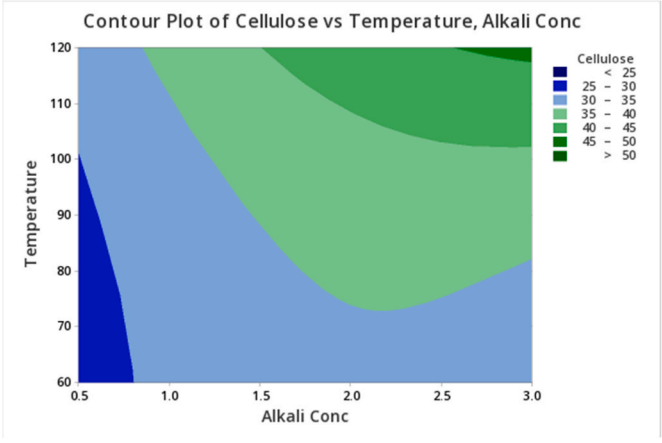


Fig. (A.2). Contour plot demonstrates the effect of alkali conc. And temperature on cellulose content. With an increase in the alkali conc. with temperature, the cellulose content also rises.

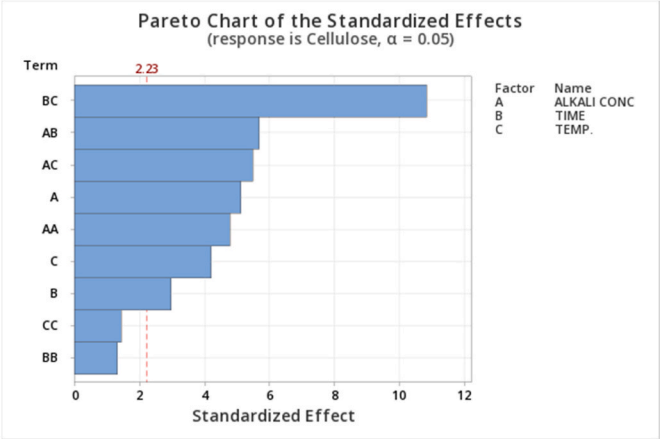


Fig. (A.3). Pareto chart of standard effects of independent variables on response. This chart illustrates that the factors time × temp., alkali conc.*time, alkali conc. *temp, alkali conc., alkali conc.*alkali conc., temp., and time are statistically significant.

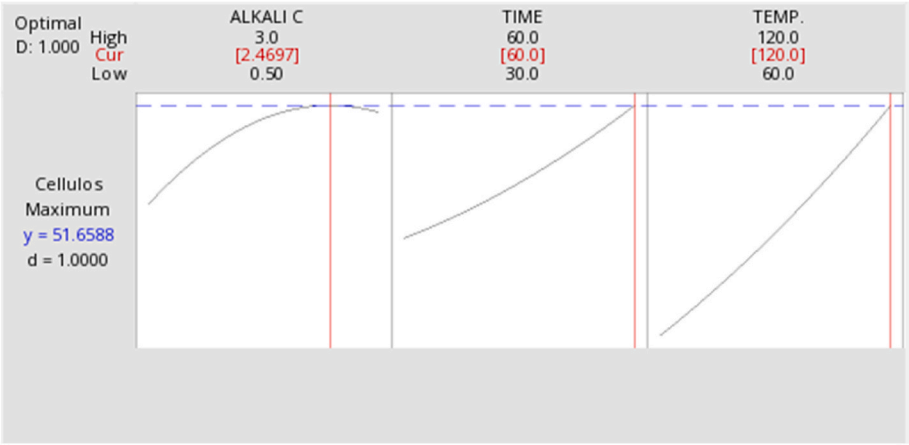


Fig. (A.4). Optimized condition demonstrates when the cellulose substrate is treated with high conc. of alkali (2.47 %) at 120 °C for 60 min then the maximum cellulose content was obtained with optimal density 1 which is statistically significant.

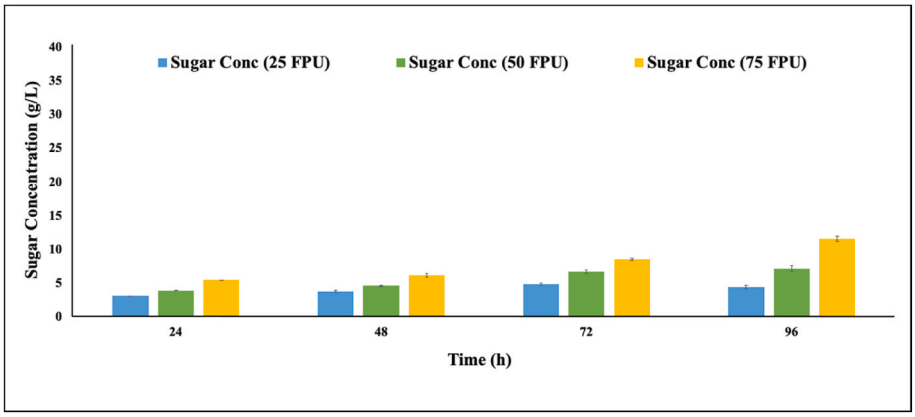


Fig. (B.1). Profile of reducing sugar concentration at the time of enzymatic hydrolysis of the raw sample of *P. stratiotes* at different initial enzyme loading. (50 °C, pH 4.8, rpm 150 with 5 % biomass loading).

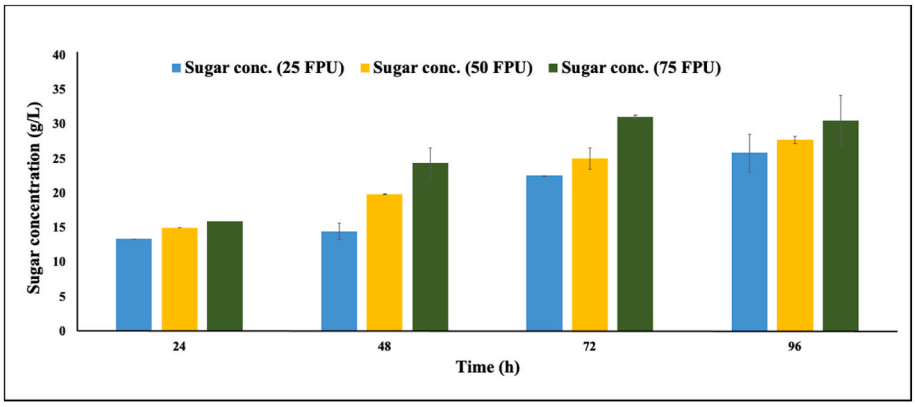


Fig. (B.2). The profile of reducing sugar concentration at the enzymatic hydrolysis optimized pre-treated sample at different enzyme loading conditions (50 °C, pH 4.8, 150 rpm and 5 % pre-treated solid content).

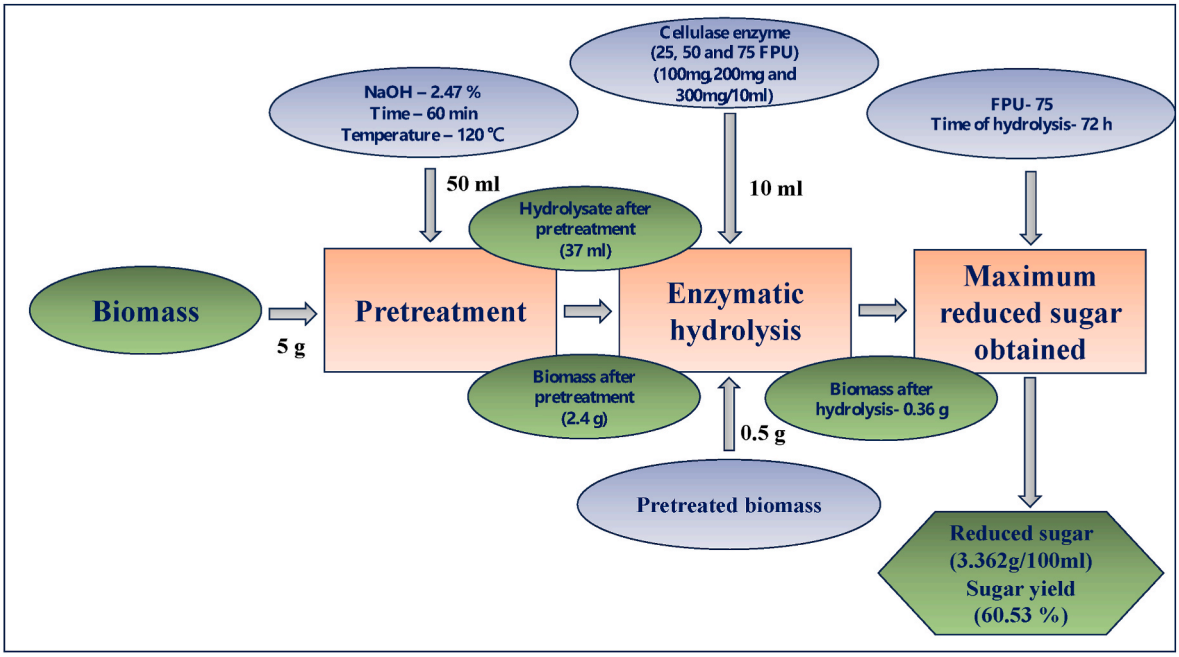


Fig. (C). The mass balance diagram of the complete process from pretreatment to enzymatic hydrolysis.

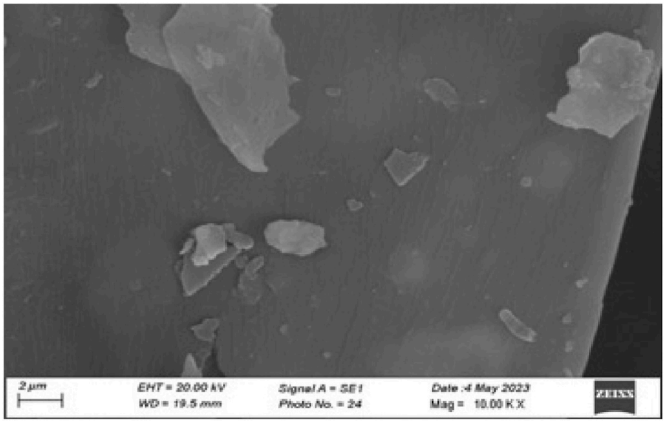


Fig. (D.1). SEM analysis for untreated (Raw).

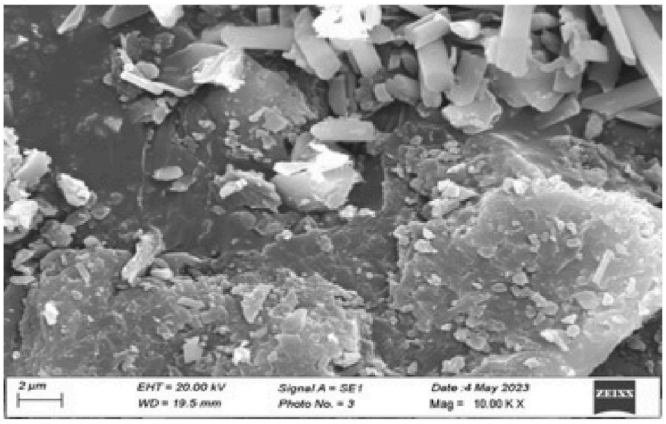


Fig. (D.2). Optimized condition (2.47 % NaOH conc., 60 min and 120 °C).

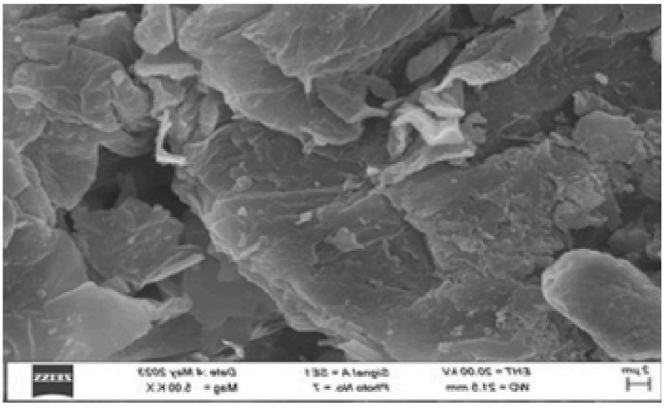


Fig. (D.3). Mild condition (0.5 % NaOH conc., 30 min and 60 °C).

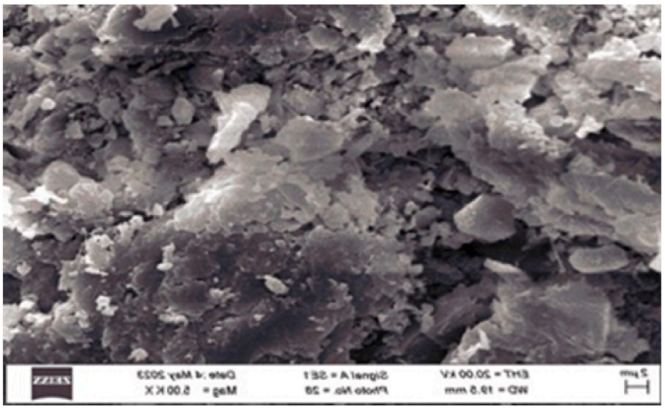


Fig. (D.4). harsh (3 % NaOH Conc., 60 min And 120 °C). Optimized condition shows maximum disruption due to the removal of lignin and hemicellulose.

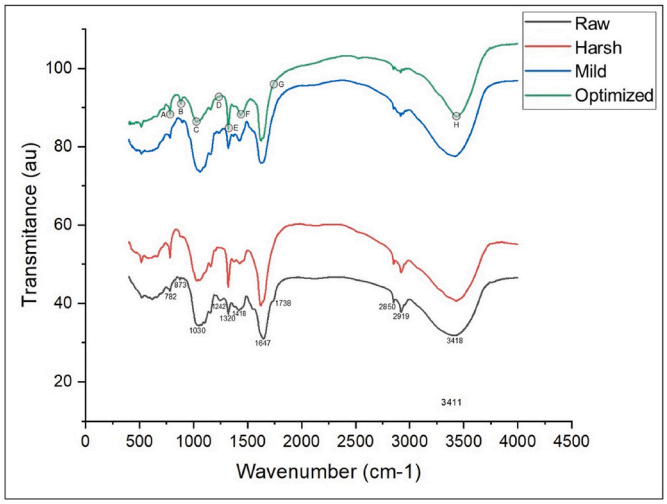


Fig. (E.1). FTIR spectra for untreated (Raw), mild, harsh and optimized conditions. The intensity of the signal in the form of peaks in different wavelengths shows the presence of relevant molecular bonds found in the polymers of samples.

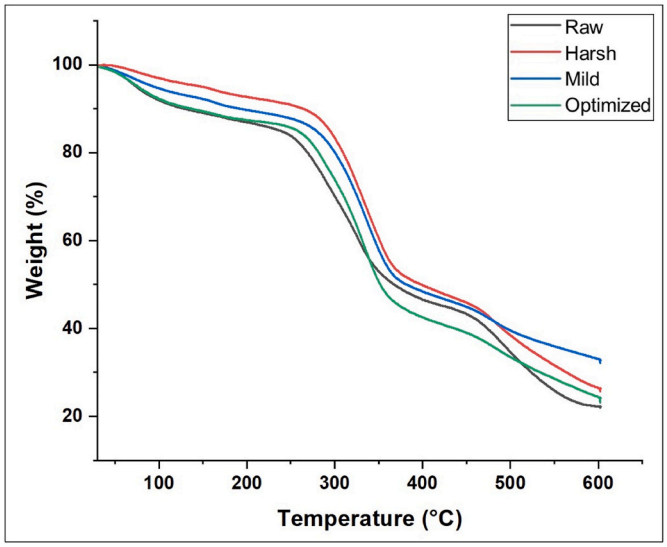


Fig. (E.2). Thermal profile for untreated (Raw), mild, harsh and optimized condition thermogravimetric analysis (TGA) curves represent the thermal degradation of the areas of the major compounds present in the samples.

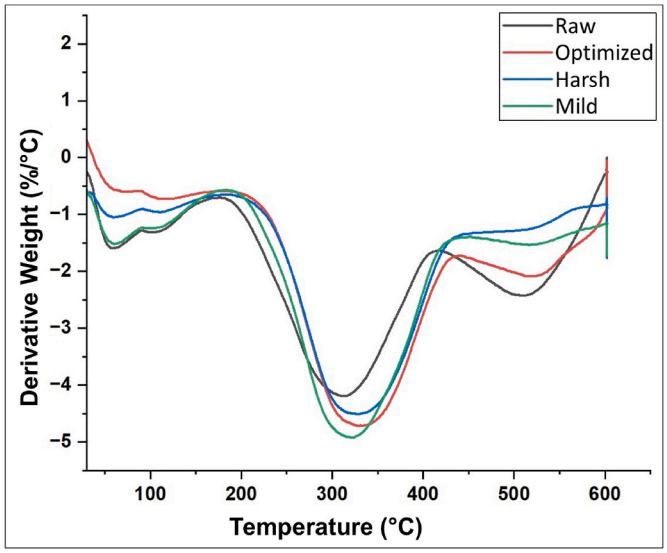


Fig. (E.3). Differential thermogravimetric (DTG) curves represent the rate of thermal degradation of treated and untreated samples.

Table (A.1)
Independent variables with experimental levels in the RSM model

Symbols	Independent variables	Unit	Relation of codes with original independent variables		
			−1	0	1
A	NaOH concentration	°C	0.5	1.75	3
B	Reaction Time	min	30	45	60
C	Reaction Temperature	% w/v	60	90	120

Table (A.2)
Experiment list with different levels of independent variables (alkali conc., time and temperature) with response (cellulose)

Experiment no.	Alkali concentration	Time	Temperature	cellulose (%)
1	1.75	45	90	37.3 ± 2.61
2	1.75	45	120	41.5 ± 3.51
3	1.75	45	90	36.92 ± 2.44
4	3	60	120	50.88 ± 1.40
5	0.5	30	60	27.38 ± 0.93
6	1.75	45	90	35.01 ± 2.95
7	3	30	60	36.31 ± 3.24
8	0.5	60	60	27.64 ± 2.61
9	0.5	30	120	21.17 ± 1.04
10	1.75	45	90	34.39 ± 2.1
11	1.75	45	60	34.2 ± 1.38
12	1.75	45	90	35.27 ± 0.72
13	3	60	60	27 ± 3.83
14	3	45	90	32.57 ± 1.91
15	3	30	120	41.12 ± 2.98
16	1.75	45	90	34.08 ± 2.51
17	1.75	60	90	40.83 ± 3.25
18	1.75	30	90	34.67 ± 2.34
19	0.5	60	120	42.28 ± 3.41
20	0.5	45	90	29.15 ± 2.39

Table (B)
Analysis of Variance (ANOVA) results and statistical parameters of the model quadratic correlation versus alkaline conc., reaction time and reaction temperature

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	9	818.693	90.966	53.79	0.000
Model	9	79.187	26.396	15.61	0.000
Linear	3	44.593	44.593	26.37	**0.000**
Alkali conc.	1	15.012	15.012	8.88	**0.014**
Time	1	30.042	30.042	17.77	**0.002**
Temp.	1	40.584	13.528	8.00	**0.005**
Square	3	38.954	38.954	23.04	**0.001**
Alkali conc.*Alkali conc.	1	3.007	3.007	1.78	0.212
Time × Time	1	3.576	3.576	2.11	0.177
Temp.*Temp.	1	305.185	101.728	60.16	**0.000**
2-Way Interaction	3	54.702	54.702	32.35	**0.000**
Alkali conc.*Time	1	51.288	51.288	30.33	**0.000**
Alkali conc.*Temp.	1	199.195	199.195	117.80	**0.000**
Time × Temp.	1	16.910	1.691		
Error	10	8.084	1.617	0.92	0.537
Lack-of-Fit	5	8.826	1.765		
Pure Error	5	835.603			
Total	19				

References

Abushammala, H., Mao, J., 2020. A review on the partial and complete dissolution and fractionation of wood and lignocelluloses using imidazolium ionic liquids. Polymers 12, 195. https://doi.org/10.3390/polym12010195.

Aguilar-Reynosa, A., Romaní, A., Ma Rodríguez-Jasso, R., Aguilar, C.N., Garrote, G., Ruiz, H.A., 2017. Microwave heating processing as alternative of pretreatment in second-generation biorefinery: an overview. Energy Convers. Manag. 136, 50–65. https://doi.org/10.1016/j.enconman.2017.01.004.

Alam, S.N., Khalid, Z., Guldhe, A., Singh, B., Korstad, J., 2021. Harvesting and pretreatment techniques of aquatic macrophytes and macroalgae for production of biofuels. Environ. Sustain 4, 299–316. https://doi.org/10.1007/S42398-021-00178-6.

Ali, C.H., Mbadinga, S.M., Liu, J.-F., Yang, S.-Z., Gu, J.-D., Mu, B.-Z., 2015. Significant enhancement of Pseudomonas aeruginosa FW_SH-1 lipase production using response surface methodology and analysis of its hydrolysis capability. J. Taiwan Inst. Chem. Eng. 52, 7–13. https://doi.org/10.1016/j.jtice.2015.02.001.

Alvira, P., Tomás-Pejó, E., Ballesteros, M., Negro, M.J., 2010. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. Bioresour. Technol. 101, 4851–4861. https://doi.org/10.1016/J.BIORTECH.2009.11.093.

Asghar, U., Irfan, M., Iram, M., Huma, Z., Nelofer, R., Nadeem, M., Syed, Q., 2015. Effect of alkaline pretreatment on delignification of wheat straw. Nat. Prod. Res. 29, 125–131. https://doi.org/10.1080/14786419.2014.964712.

Awasthi, M., Kaur, J., Rana, S., 2013. Bioethanol production through water hyacinth, Eichhornia crassipes via optimization of the pretreatment conditions. Int. J. Emerg. Technol. Adv. Eng. 3, 42–46.

Bano, A., Irfan, M., 2019. Alkali pretreatment of cotton stalk for bioethanol. Bangladesh J. Sci. Ind. Res. 54, 73–82. https://doi.org/10.3329/BJSIR.V54I1.40733.

Barua, V.B., Kalamdhad, A.S., 2017. Effect of various types of thermal pretreatment techniques on the hydrolysis, compositional analysis and characterization of water hyacinth. Bioresour. Technol. 227, 147–154. https://doi.org/10.1016/J.BIORTECH.2016.12.036.

Baruah, J., Deka, R., Kalita, E., 2020. Greener production of microcrystalline cellulose (MCC) from Saccharum spontaneum (Kans grass): statistical optimization. Int. J. Biol. Macromol. 154, 672–682.

Basak, R.K., Saha, S.G., Sarkar, A.K., Saha, M., Das, N.N., Mukherjee, A.K., 1993. Thermal properties of jute constituents and flame retardant jute fabrics. Textil. Res. J. 63, 658–666. https://doi.org/10.1177/004051759306301107.

Başar, İ.A., Çoban, Ö., Göksungur, M.Y., Eskicioğlu, Ç., Perendeci, N.A., 2021. Enhancement of lignocellulosic biomass anaerobic digestion by optimized mild alkaline hydrogen peroxide pretreatment for biorefinery applications. J. Environ. Manag. 298, 113539 https://doi.org/10.1016/J.JENVMAN.2021.113539.

Bhati, N., Sharma, A.K., 2023. Comparative study of different chemical pretreatments for enhanced enzymatic hydrolysis of sorghum straw. Biomass Convers. Biorefinery 1, 1–9. https://doi.org/10.1007/S13399-023-05185-7/FIGURES/4.

Chen, K.Y., Zheng, Y., Cheng, Y.S., 2015. Integrated alkali pretreatment and preservation of wet lettuce (Pistia stratiotes) by lactic acid bacteria for fermentable sugar production. Biomass Bioenergy 81, 249–255. https://doi.org/10.1016/J.BIOMBIOE.2015.07.007.

Culaba, A.B., Mayol, A.P., San Juan, J.L.G., Vinoya, C.L., Concepcion, R.S., Bandala, A. A., Vicerra, R.R.P., Ubando, A.T., Chen, W.H., Chang, J.S., 2022. Smart sustainable biorefineries for lignocellulosic biomass. Bioresour. Technol. 344, 126215 https://doi.org/10.1016/J.BIORTECH.2021.126215.

Dahunsi, S.O., Adesulu-Dahunsi, A.T., Izebere, J.O., 2019. Cleaner energy through liquefaction of Cocoa (Theobroma cacao) pod husk: pretreatment and process optimization. J. Clean. Prod. 226, 578–588. https://doi.org/10.1016/j.jclepro.2019.04.112.

Deng, Z., Xia, A., Liao, Q., Zhu, X., Huang, Y., Fu, Q., 2019. Laccase pretreatment of wheat straw: effects of the physicochemical characteristics and the kinetics of enzymatic hydrolysis. Biotechnol. Biofuels 12. https://doi.org/10.1186/S13068-019-1499-3.

Dessie, W., Xiao, J., Tang, J., An, B., Luo, X., Wang, M., Liao, Y., Wahab, R., Li, C., Qin, Z., 2024. Maximizing fermentable feedstocks from Camellia oleifera seed oil extraction residues: green pretreatment and enzymatic hydrolysis for effective valorization. Arab. J. Chem. 17, 105815 https://doi.org/10.1016/j.arabjc.2024.105815.

Fillat, Ú., Ibarra, D., Eugenio, M.E., Moreno, A.D., Tomás-Pejó, E., Martín-Sampedro, R., 2017. Laccases as a potential tool for the efficient conversion of lignocellulosic biomass: a review, 2017 Fermentatio 3. https://doi.org/10.3390/FERMENTATION3020017. Page 17 3, 17.

Fontes, D.I., Bezerra, T.S., de Freitas, E.P.B., de Oliveira, M.N., Silva, S. da C., Silva, S.Y. S., Albino, U.B., Santos, D. de A., 2023. Production of cellulases from Amazonian fungi and their application in babassu cellulose hydrolysis. Int. Biodeterior. Biodegrad. 182, 105631 https://doi.org/10.1016/J.IBIOD.2023.105631.

Gao, L., Gu, J.D., 2021. A new unified conceptual framework involving maintenance energy, metabolism and toxicity for research on degradation of organic pollutants. Int. Biodeterior. Biodegrad. 162, 105253 https://doi.org/10.1016/J.IBIOD.2021.105253.

Gu, J.D., 2020. On environmental biotechnology of bioremediation. Appl. Environ. Biotechnol. 5, 3–8. https://doi.org/10.26789/AEB.2020.02.002.

Gu, J.D., 2016. Biodegradation testing: so many tests but very little new innovation. Appl. Environ. Biotechnol. 1, 92–95. https://doi.org/10.26789/AEB.2016.01.007.

Gundupalli, M.P., Tantayotai, P., Panakkal, E.J., Chuetor, S., Kirdponpattara, S., Thomas, A.S.S., Sharma, B.K., Sriariyanun, M., 2022. Hydrothermal pretreatment optimization and deep eutectic solvent pretreatment of lignocellulosic biomass: an integrated approach. Bioresour. Technol. Rep. 17, 100957 https://doi.org/10.1016/J.BITEB.2022.100957.

Gusain, R., Suthar, S., 2017. Potential of aquatic weeds (Lemna gibba, Lemna minor, Pistia stratiotes and Eichhornia sp.) in biofuel production. Process Saf. Environ. Protect. 109, 233–241. https://doi.org/10.1016/J.PSEP.2017.03.030.

Hamann, P.R.V., Serpa, D.L., Barreto da Cunha, A.S., de Camargo, B.R., Osiro, K.O., Valle de Sousa, M., Felix, C.R., Miller, R.N.G., Noronha, E.F., 2015. Evaluation of plant cell wall degrading enzyme production by Clostridium thermocellum B8 in the presence of raw agricultural wastes. Int. Biodeterior. Biodegrad. 105, 97–105. https://doi.org/10.1016/J.IBIOD.2015.08.013.

Hernández-Mendoza, A.G., Saldaña-Trinidad, S., Martínez-Hernández, S., Pérez-Sariñana, B.Y., Láinez, M., 2021. Optimization of alkaline pretreatment and enzymatic hydrolysis of cocoa pod husk (Theobroma cacao L.) for ethanol production. Biomass Bioenergy 154, 106268. https://doi.org/10.1016/J.BIOMBIOE.2021.106268.

Holtzapple, M.T., 2003. In: Caballero, B.B.T.-E., of, F.S. and N. (Eds.), HEMICELLULOSES, Second E. Academic Press, Oxford, pp. 3060–3071. https://doi.org/10.1016/B0-12-227055-X/00589-7.

Hu, Y., Priya, A., Chen, C., Liang, C., Wang, W., Wang, Q., Lin, C.S.K., Qi, W., 2023a. Recent advances in substrate-enzyme interactions facilitating efficient biodegradation of lignocellulosic biomass: a review. Int. Biodeterior. Biodegrad. 180 https://doi.org/10.1016/j.ibiod.2023.105594.

Hu, Y., Priya, A., Chen, C., Liang, C., Wang, W., Wang, Q., Lin, C.S.K., Qi, W., 2023b. Recent advances in substrate-enzyme interactions facilitating efficient biodegradation of lignocellulosic biomass: a review. Int. Biodeterior. Biodegrad. 180, 105594 https://doi.org/10.1016/J.IBIOD.2023.105594.

Infanzón-Rodríguez, M.I., Ragazzo-Sánchez, J.A., del Moral, S., Calderón-Santoyo, M., Aguilar-Uscanga, M.G., 2022. Enzymatic hydrolysis of lignocellulosic biomass using native cellulase produced by Aspergillus Niger ITV02 under liquid state fermentation. Biotechnol. Appl. Biochem. 69, 198–208. https://doi.org/10.1002/BAB.2097.

Jayanth, K.P., 2000. Biological Control of Weeds in India, Biocontrol Potential and its Exploitation in Sustainable Agriculture. CSIRO Publishing. https://doi.org/10.1007/978-1-4615-4209-4_15.

Ji, G., Han, L., Gao, C., Xiao, W., Zhang, Y., Cao, Y., 2017. Quantitative approaches for illustrating correlations among the mechanical fragmentation scales, crystallinity and enzymatic hydrolysis glucose yield of rice straw. Bioresour. Technol. 241, 262–268. https://doi.org/10.1016/j.biortech.2017.05.062.

Jin, S., Zhang, G., Zhang, P., Jin, L., Fan, S., Li, F., 2015. Comparative study of high-pressure homogenization and alkaline-heat pretreatments for enhancing enzymatic hydrolysis and biogas production of grass clipping. Int. Biodeterior. Biodegrad. 104, 477–481. https://doi.org/10.1016/j.ibiod.2015.08.005.

Kataria, R., Mol, A., Schulten, E., Happel, A., Mussatto, S.I., 2017. Bench scale steam explosion pretreatment of acid impregnated elephant grass biomass and its impacts on biomass composition, structure and hydrolysis. Ind. Crops Prod. 106, 48–58. https://doi.org/10.1016/J.INDCROP.2016.08.050.

Kataria, R., Woods, T., Casey, W., Cerrone, F., Davis, R., O'Connor, K., Ruhal, R., Babu, R., 2018a. Surfactant-mediated hydrothermal pretreatment of Ryegrass followed by enzymatic saccharification for polyhydroxyalkanoate production. Ind. Crops Prod. 111, 625–632. https://doi.org/10.1016/J.INDCROP.2017.11.029.

Kataria, R., Woods, T., Casey, W., Cerrone, F., Davis, R., O'Connor, K., Ruhal, R., Babu, R., 2018b. Surfactant-mediated hydrothermal pretreatment of Ryegrass followed by enzymatic saccharification for polyhydroxyalkanoate production. Ind. Crops Prod. 111, 625–632. https://doi.org/10.1016/J.INDCROP.2017.11.029.

Kian, L., Jawaid, M., Ariffin, H., Alothman, O., 2017. Isolation and characterization of microcrystalline cellulose from roselle fibers. Int. J. Biol. Macromol. 103, 931–940.

Kim, I., Han, J.I., 2012. Optimization of alkaline pretreatment conditions for enhancing glucose yield of rice straw by response surface methodology. Biomass Bioenergy 46, 210–217. https://doi.org/10.1016/J.BIOMBIOE.2012.08.024.

Kumar, M., Turner, S., 2015. Protocol: a medium-throughput method for determination of cellulose content from single stem pieces of Arabidopsis thaliana. Plant Methods 11. https://doi.org/10.1186/S13007-015-0090-6.

Li, F., Zhang, P., Zhang, G., Tang, X., Wang, S., Jin, S., 2017. Enhancement of corn stover hydrolysis with rumen fluid pretreatment at different solid contents: effect, structural changes and enzymes participation. Int. Biodeterior. Biodegrad. 119, 405–412. https://doi.org/10.1016/j.ibiod.2016.10.038.

Li, X., Shi, Y., Kong, W., Wei, J., Song, W., Wang, S., 2022. Improving enzymatic hydrolysis of lignocellulosic biomass by bio-coordinated physicochemical pretreatment—a review. Energy Rep. https://doi.org/10.1016/j.egyr.2021.12.015.

Liu, P., Li, A., Wang, Youmei, Cai, Q., Yu, H., Li, Y., Peng, H., Li, Q., Wang, Yanting, Wei, X., Zhang, R., Tu, Y., Xia, T., Peng, L., 2021. Distinct *Miscanthus* lignocellulose improves fungus secreting cellulases and xylanases for consistently enhanced biomass saccharification of diverse bioenergy crops. Renew. Energy 174, 799–809. https://doi.org/10.1016/J.RENENE.2021.04.107.

Lu, P., Hsieh, Y. Lo, 2010. Preparation and properties of cellulose nanocrystals: rods, spheres, and network. Carbohydr. Polym. 82, 329–336. https://doi.org/10.1016/j.carbpol.2010.04.073.

Malik, K., Salama, E.S., Kim, T.H., Li, X., 2020. Enhanced ethanol production by Saccharomyces cerevisiae fermentation post acidic and alkali chemical pretreatments of cotton stalk lignocellulose. Int. Biodeterior. Biodegrad. 147 https://doi.org/10.1016/j.ibiod.2019.104869.

Mankar, A.R., Pandey, A., Modak, A., Pant, K.K., 2021. Pretreatment of lignocellulosic biomass: a review on recent advances. Bioresour. Technol. 334, 125235 https://doi.org/10.1016/J.BIORTECH.2021.125235.

Manmai, N., Unpaprom, Y., Ponnusamy, V.K., Ramaraj, R., 2020. Bioethanol production from the comparison between optimization of sorghum stalk and sugarcane leaf for sugar production by chemical pretreatment and enzymatic degradation. Fuel 278. https://doi.org/10.1016/j.fuel.2020.118262.

Meehnian, H., Jana, A.K., Jana, M.M., 2017. Pretreatment of cotton stalks by synergistic interaction of Daedalea flavida and Phlebia radiata in co-culture for improvement in delignification and saccharification. Int. Biodeterior. Biodegrad. 117, 68–77. https://doi.org/10.1016/J.IBIOD.2016.11.022.

Miller, G.L., 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. 31, 426–428. https://doi.org/10.1021/AC60147A030.

Mishima, D., Kuniki, M., Sei, K., Soda, S., Ike, M., Fujita, M., 2008. Ethanol production from candidate energy crops: water hyacinth (Eichhornia crassipes) and water lettuce (Pistia stratiotes L.). Bioresour. Technol. 99, 2495–2500. https://doi.org/10.1016/J.BIORTECH.2007.04.056.

Moerman, D.E., 1996. An analysis of the food plants and drug plants of native North America. J. Ethnopharmacol. 52, 1–22. https://doi.org/10.1016/0378-8741(96)01393-1.

Nazari, M.T., Mazutti, J., Basso, L.G., Colla, L.M., Brandli, L., 2021. Biofuels and their connections with the sustainable development goals: a bibliometric and systematic review. Environ. Dev. Sustain. 23, 11139–11156. https://doi.org/10.1007/S10668-020-01110-4.

Njoku, S.I., Ahring, B.K., Uellendahl, H., 2013. Tailoring wet explosion process parameters for the pretreatment of cocksfoot grass for high sugar yields. Appl. Biochem. Biotechnol. 170, 1574–1588. https://doi.org/10.1007/S12010-013-0299-7.

Nomanbhay, S.M., Hussain, R., Palanisamy, K., Nomanbhay, S.M., Hussain, R., Palanisamy, K., 2013. Microwave-assisted alkaline pretreatment and microwave assisted enzymatic saccharification of oil palm empty fruit bunch fiber for enhanced fermentable sugar yield. J. Sustain. Bioenergy Syst. 3, 7–17. https://doi.org/10.4236/JSBS.2013.31002.

Pant, S., Ritika, Komesu, A., Penteado, E.D., Diniz, A.A.R., Rahman, M.A., Kuila, A., 2021. NaOH pretreatment and enzymatic hydrolysis of Brassica juncea using mixture of cellulases. Environ. Technol. Innov. 21, 101324 https://doi.org/10.1016/J.ETI.2020.101324.

Pantawong, R., Chuanchai, A., Thipbunrat, P., Unpaprom, Y.R.R., 2015. Experimental investigation of biogas production from water lettuce, Pistia stratiotes L. Emer. Life Sci. Res. 1, 41–46.

Pashaei, H., Ghaemi, A., Nasiri, M., Karami, B., 2020. Experimental modeling and optimization of CO2 absorption into piperazine solutions using RSM-CCD methodology. ACS Omega 5, 8432–8448. https://doi.org/10.1021/acsomega.9b03363.

Punia, P., Singh, L., 2024. Optimization of alkali pre-treatment of sweet sorghum [Sorghum bicolor (L.) Moench] residue to improve enzymatic hydrolysis for fermentable sugars. Waste Manag. Bull. 2, 131–141. https://doi.org/10.1016/j.wmb.2023.12.007.

Ramaraj, R., Unpaprom, Y., 2019. Optimization of pretreatment condition for ethanol production from Cyperus difformis by response surface methodology. 3 Biotech 9. https://doi.org/10.1007/S13205-019-1754-0.

Reddy, K.O., Maheswari, C.U., Dhlamini, M.S., Mothudi, B.M., Kommula, V.P., Zhang, Jinming, Zhang, Jun, Rajulu, A.V., 2018. Extraction and characterization of cellulose single fibers from native african napier grass. Carbohydr. Polym. 188, 85–91. https://doi.org/10.1016/j.carbpol.2018.01.110.

Rezania, S., Oryani, B., Cho, J., Talaiekhozani, A., Sabbagh, F., Hashemi, B., Rupani, P.F., Mohammadi, A.A., 2020. Different pretreatment technologies of lignocellulosic biomass for bioethanol production: an overview. Energy 199. https://doi.org/10.1016/j.energy.2020.117457.

Rodrigues Gurgel da Silva, A., Giuliano, A., Errico, M., Rong, B.G., Barletta, D., 2019. Economic value and environmental impact analysis of lignocellulosic ethanol production: assessment of different pretreatment processes. Clean Technol. Environ. Policy 21, 637–654. https://doi.org/10.1007/s10098-018-01663-z.

Saha, B.C., Qureshi, N., Kennedy, G.J., Cotta, M.A., 2016. Biological pretreatment of corn stover with white-rot fungus for improved enzymatic hydrolysis. Int. Biodeterior. Biodegrad. 109, 29–35. https://doi.org/10.1016/j.ibiod.2015.12.020.

Selvakumar, P., Adane, A.A., Zelalem, T., Hunegnaw, B.M., Karthik, V., Kavitha, S., Jayakumar, M., Karmegam, N., Govarthanan, M., Kim, W., 2022. Optimization of binary acids pretreatment of corncob biomass for enhanced recovery of cellulose to produce bioethanol. Fuel 321. https://doi.org/10.1016/j.fuel.2022.124060.

Sills, D.L., Gossett, J.M., 2012. Using FTIR spectroscopy to model alkaline pretreatment and enzymatic saccharification of six lignocellulosic biomasses. Biotechnol. Bioeng. 109, 894–903. https://doi.org/10.1002/BIT.24376.

Sombatpraiwan, S., Junyusen, T., Treeamnak, T., Junyusen, P., 2019. Optimization of microwave-assisted alkali pretreatment of cassava rhizome for enhanced enzymatic hydrolysis glucose yield. Food Energy Secur. 8 https://doi.org/10.1002/FES3.174.

Sulyman, A.O., Igunnu, A., Malomo, S.O., 2020. Isolation, purification and characterization of cellulase produced by Aspergillus Niger cultured on Arachis hypogaea shells. Heliyon 6, e05668. https://doi.org/10.1016/J.HELIYON.2020.E05668.

Suman, S.K., Dhawaria, M., Tripathi, D., Raturi, V., Adhikari, D.K., Kanaujia, P.K., 2016. Investigation of lignin biodegradation by Trabulsiella sp. isolated from termite gut. Int. Biodeterior. Biodegrad. 112, 12–17. https://doi.org/10.1016/j.ibiod.2016.04.036.

Trevorah, R.M., Othman, M.Z., 2015. Alkali pretreatment and enzymatic hydrolysis of Australian timber mill sawdust for biofuel production. J. Renew. Energy 2015, 1–9. https://doi.org/10.1155/2015/284250.

Tse, T.J., Wiens, D.J., Reaney, M.J.T., 2021. Production of bioethanol—a review of factors affecting ethanol yield, 2021 Fermentatio 7. https://doi.org/10.3390/FERMENTATION7040268, 268 7, 268.

Umesh, M., Santhosh, A.S., Shanmugam, S., Thazeem, B., Alharbi, S.A., Almoallim, H.S., Chi, N.T.L., Pugazhendhi, A., 2022. Extraction, characterization, and fabrication of cellulose biopolymer sheets from Pistia stratiotes as a biodegradative coating material: an unique strategy for the conversion of invasive weeds into value-added products. J. Polym. Environ. 30, 5057–5068. https://doi.org/10.1007/S10924-022-02511-4/FIGURES/9.

Valles, A., Capilla, M., Álvarez-Hornos, F.J., García-Puchol, M., San-Valero, P., Gabaldón, C., 2021. Optimization of alkali pretreatment to enhance rice straw conversion to butanol. Biomass Bioenergy 150, 106131. https://doi.org/10.1016/J.BIOMBIOE.2021.106131.

Wannapokin, A., Ramaraj, R., Whangchai, K., Unpaprom, Y., 2018. Potential improvement of biogas production from fallen teak leaves with co-digestion of microalgae. 3 Biotech 8. https://doi.org/10.1007/S13205-018-1084-7.

Wu, H., Zhao, X., Adsul, M., Zhong, Y., Qian, Y., Zhong, L., Sun, Y., Sun, N., Zhang, L., Liu, W., Qu, Y., 2019. Enhancement of cellulase production in Trichoderma reesei via disruption of multiple protease genes identified by comparative secretomics. Enhancement of Cellulase Production in Trichoderma reesei via Disruption of Multiple Protease Genes Identified by Comparative Secretomics 10, 2784. https://doi.org/10.3389/fmicb.2019.02784.

Xia, M., Valverde-Barrantes, O.J., Suseela, V., Blackwood, C.B., Tharayil, N., 2022. Characterizing natural variability of lignin abundance and composition in fine roots across temperate trees: a comparison of analytical methods. New Phytol. https://doi.org/10.1111/NPH.18515.

Yadav, D., Barbora, L., Bora, D., Mitra, S., Rangan, L., Mahanta, P., 2017. An assessment of duckweed as a potential lignocellulosic feedstock for biogas production. Int. Biodeterior. Biodegrad. 119, 253–259. https://doi.org/10.1016/J.IBIOD.2016.09.007.

Yang, M., Zhang, Xiaoliang, Wang, K., Zhu, S., Ye, Z., Sheng, K., Zhang, Ximing, 2022. Investigation of cascade valorization of Pistia stratiotes L. by hydrothermal treatment. Fuel 324, 124473. https://doi.org/10.1016/J.FUEL.2022.124473.

Zhou, Y., Zheng, J., Gan, R.-Y., Zhou, T., Xu, D.-P., Li, H.-B., Cravotto, G., Choi, Y.H., 2017. Optimization of ultrasound-assisted extraction of antioxidants from the mung bean coat. Molecules. https://doi.org/10.3390/molecules22040638.

Check for updates

Received: 26 March 2024 | Revised: 3 June 2024 | Accepted: 27 June 2024

DOI: 10.1002/suco.202400387

ARTICLE

fib WILEY

Enhancing concrete durability and strength: An innovative approach integrating abrasion and cement slurry treatment for recycled coarse aggregates

Harish Panghal | **Awadhesh Kumar**

Department of Civil Engineering, Delhi Technological University, Delhi, India

Correspondence
Harish Panghal, Department of Civil Engineering, Delhi Technological University, Delhi 1100042, India.
Email: harish_phd2k18@dtu.ac.in

Abstract

This study introduces a novel approach to enhance the durability and strength of concrete by integrating abrasion and cement slurry treatments on surface-treated recycled coarse aggregates (STRCA). Initial abrasion treatment removes aged mortar from the recycled coarse aggregate (RCA) surface, while subsequent cement slurry modification provides a fresh surface, reinforcing the bond with the concrete matrix. Through detailed material analyses including x-ray diffraction, scanning electron microscopy, and energy-dispersive x-ray, microstructural changes are evaluated alongside compressive strength, drying shrinkage, electrical resistivity, and chloride ion penetration of STRCA-based concrete. Results indicate significant improvements in mechanical properties and durability as the treatments eliminate mortar and enhance the interfacial connection with cement paste. Notably, a substantial 29.37% strength enhancement is observed at 25% replacement (STRCA 25), demonstrating treatment efficacy. While a reduction of 11.58% occurs at 100% replacement (STRCA 100), up to 75% replacement shows minimal strength loss, with optimal enhancements at 50% replacement. Additionally, STRCA concrete exhibits lower drying shrinkage (11%–20% reduction) and increased electrical resistivity (28%–36% rise), indicating improved durability at 100% replacement. Enhanced resistance to chloride ion penetration is also evident, with a 22%–28% increase at 100% replacement compared to conventional RCA. Overall, this study highlights the effectiveness of dual treatments for enhancing concrete properties using STRCA.

KEYWORDS

abrasion treatment, cement slurry treatment, concrete durability, microstructural characteristics, surface-treated recycled coarse aggregates

1 | INTRODUCTION

The confluence of economic expansion and urbanization has emerged as a driving force behind environmental degradation,[1] posing a formidable threat to sustainability.[2] The escalating demand for natural resources, particularly within the construction sector, has given rise to a significant volume of construction waste, often inadequately managed in landfills.[3,4] This pervasive waste, spanning diverse sectors, presents complex environmental

 © 2024 *fib*. International Federation for Structural Concrete

challenges.[5] Recycling emerges as a pivotal solution, offering a multitude of benefits such as resource conservation, diminished CO_2 emissions, and bolstered economic competitiveness.[6–8] The concrete industry, notably cement, is highlighted as having a significant contribution to environmental harm and carbon emissions.[9–11] Recycling construction materials not only addresses environmental concerns but also bestows economic advantages by extracting value from waste, reducing environmental protection costs, and curbing waste transportation expenses.[12–15] In addressing environmental considerations, numerous studies have been conducted to advocate for the integration of recycled aggregates in the construction of concrete.[16–20] Many studies have examined the mechanical properties and durability of concretes containing recycled coarse aggregate (RCA), focusing on aspects like water permeability, chloride ion permeability, corrosion resistance, and resistance to carbonation.[21–24] Nevertheless, apprehensions have arisen regarding diminished mechanical strength and durability, attributed to factors like reduced density, increased water absorption, and the presence of mortar adhering to recycled particles.[25–29]

Numerous studies have been undertaken to enhance the performance of RCA, employing a diverse array of methods. These methods include the removal of old mortar using acid solutions[30–32] and abrasion machines,[33–35] as well as fortifying the old mortar through processes such as CO_2 treatment,[36–38] cement-based slurries,[39,40] nanomaterials,[41–43] sodium silicate,[44] and bacteria,[45] among others. While these approaches have shown promise in improving RCA performance, they are not without challenges. The use of acid solutions presents environmental risks, and the abrasion method requires additional energy, contributing to CO_2 emissions. Moreover, CO_2 and bacteria treatments often involve extended durations. Addressing these concerns, enhancing the old mortar of RCA through a solution (slurry) emerges as a promising alternative. Innovative mixing techniques, such as early incorporation of cement and recycled aggregates, show promise in enhancing compressive strength over time.[46,47] Even with the use of recycled aggregates, performance comparable to conventional concrete can be achieved by adjusting the water-to-cement ratio.[48–51] Novel mixing methods, involving the pre-treatment of recycled aggregates with specific substances, hold potential for preserving the modulus of elasticity.[52,53] Variables such as replacement content, mixing techniques, aggregate type, and porosity significantly influence the mechanical properties of recycled aggregate concrete (RAC).[54,55] Durability emerges as a critical aspect, and various tests, including electrical resistivity tests, can provide estimates in this regard.[56,57] In general, RAC exhibits lower electrical resistance than regular concrete, with variations dependent on factors such as concrete composition, environmental conditions, and exposure to aggressive agents.[58–60]

This research aims to address critical challenges and bridge knowledge gaps related to the integration of recycled aggregates in concrete manufacturing. Despite the positive environmental impact of recycled aggregates, obstacles such as impurities, mechanical vulnerabilities, and bonding difficulties pose challenges to concrete durability.[61–63] In response, various researchers have explored different treatment techniques, including cement slurry coating, abrasion treatment, chemical treatment, carbonation treatment, and various combinations of these methods, to modify the surface of recycled aggregates and enhance the bond between aggregates and the cement matrix.[64–67] Out of the different treatment solutions and slurries available, cement-based slurry stands out as a commonly employed choice due to its economic feasibility and ready availability. Importantly, the impact of applying abrasion treatment followed by cement slurry treatment on both strength and durability has not been previously studied. This study introduces a novel approach that utilizes abrasion and cement slurry treatments to improve the connection between recycled materials and the concrete matrix. The expectation is that implementing this technique will significantly enhance the durability of RAC, aligning with the principles of the circular economy. While existing studies primarily focus on mechanical properties, there is a noticeable gap in research dedicated explicitly to concrete durability. This study, however, focuses on surface-treated recycled coarse aggregates (STRCA), delving into parameters such as drying shrinkage, electrical resistivity, and chloride ion penetration. The objective is to elucidate the mechanical and physical characteristics of STRCA, quantify the bond strength with cement paste, and scrutinize the durability of concrete containing these modified aggregates. By systematically addressing these aspects, this study introduces an innovative perspective to concrete technology, providing a comprehensive framework for the effective and sustainable utilization of recycled aggregates in concrete production.

2 | EXPERIMENTAL PROGRAMS

2.1 | Materials

Ordinary Portland cement of grade 43 that complies with IS 269-2015[68] standards, served as the primary binder in this investigation for concrete. The natural fine aggregates (NFA) comprised natural sand, complying with IS

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

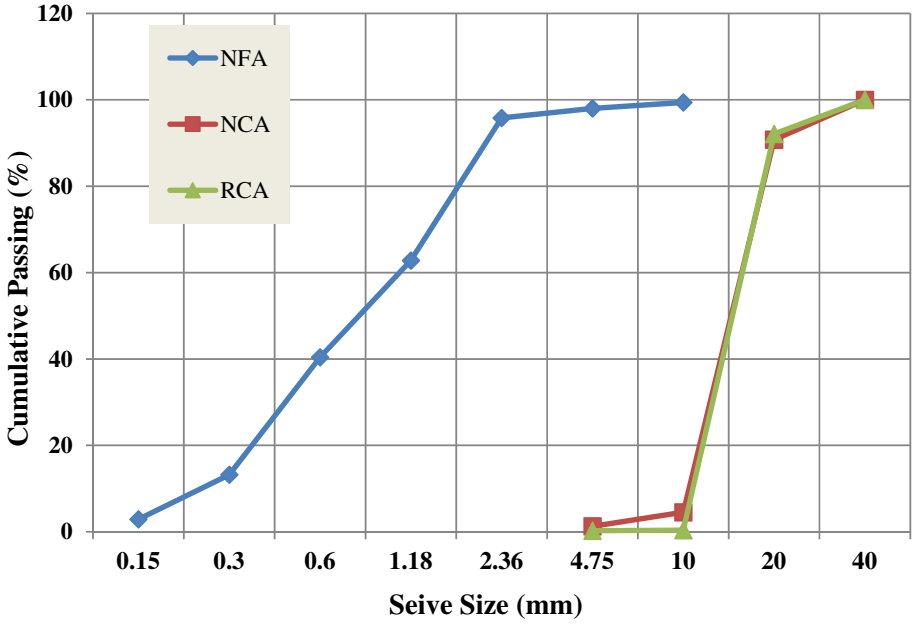


FIGURE 1 Particle size distribution curves. NCA, natural coarse aggregates; NFA, natural fine aggregates; RCA, recycled coarse aggregate.

TABLE 1 Different characteristic of aggregates.

Property	NFA	NCA	RCA	STRCA	Standard limits
Bulk density (kg/m^3)	1625	1740	1660	1704	1200–1750 (IS 2386 Part 3)[70]
Specific gravity	2.675	2.754	2.681	2.607	2.30–2.90 (IS 2386 Part 3)[70]
Water absorption (%)	0.51	1.12	5.35	5.51	≤2.0 (IS 2386 Part 3)[70]
Abrasion loss (%)	15.52	25.43	28.76	27.46	<30 (IS 2386 Part 4)[71]
Crushing value (%)	16.12	26.42	27.17	26.63	< 30 (IS 2386 Part 4)[71]
Impact value (%)	15.13	17.41	20.86	15.62	<30 (IS 2386 Part 4)[71]

Abbreviations: NCA, natural coarse aggregates, NFA, natural fine aggregates; RCA, recycled coarse aggregate; STRCA, surface-treated recycled coarse aggregates.

383-2016[69] standards, while the natural coarse aggregates (NCA) were sourced from crushed stone following the mix proportion guidelines outlined in IS 383-2016.[69] Waste from building and demolition was used to create RCA, which was processed with an impact crusher to meet specified dimensions. The distribution of aggregate particle sizes is illustrated in Figure 1. Table 1 provides a detailed presentation of the thorough physical and mechanical analyses conducted on the utilization of RCAs in concrete. A notable bonding challenge stems from the presence of residual old cement mortar on aggregate surfaces, resulting in a strength reduction. To overcome this, an integrated approach combining abrasion and cement slurry treatments is applied. Abrasion treatment yields a coarse surface, and cement slurry treatment produces a smooth one. These STRCA are commonly termed STRCA.

Aggregate specific gravity and water absorption were tested in accordance with IS 2386 Part 3.[70] The specific gravity values for NFA, NCA, and RCA are 2.675, 2.754, and 2.681, respectively, conforming to the standard limits of 2.30–2.90, indicating compliance with the specified requirements for specific gravity. In terms of water absorption, NFA exhibits the lowest absorption at 0.51%, followed by NCA at 1.12%. In contrast, RCA displays a higher water absorption of 5.35%. According to IS 9103–1999,[72] a chemical admixture—specifically, superplasticizer (C-MAX)—was used as an additive at a rate of 1% by cement weight. Mixing and curing of the materials were conducted using potable water available within the laboratory.

2.2 | Surface treatments of RCA

In the pursuit of mitigating the substantial strength reduction observed in RAC due to insufficient bonding with the cement matrix, a novel approach involving abrasion and cement slurry treatments was implemented. The aim was to enhance the surfaces of RCA, addressing the bonding challenges attributed to aged mortar residues. The comprehensive process, depicted in Figure 2,

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Los Angeles Abrasion

(b) Abrasion Treated RCA

(c) Abrasion Treated RCA Immersed in Cement Slurry

(d) Surface Treated RCA (STRCA)

FIGURE 2 Illustration of the surface-treated recycled coarse aggregates (STRCA) production process. RCA, recycled coarse aggregate.

TABLE 2 Weight loss and water absorption characteristics of recycled coarse aggregate (RCA) before and during abrasion test.

Weight of RCA Before the test (kg)	Weight of RCA after 1 min of test duration (kg)	Weight loss After 1 min test duration (%)	Weight of RCA after 2 min of test duration (kg)	Weight loss after 2 min test duration (%)	Weight of RCA after 5 min of test duration (kg)	Weight loss after 5 min test duration (%)
10	8.36	4.6	8.70	12.52	8.75	12.52
10	9.93	5.9	9.38	11.17	8.64	14.03
10	8.83	5.3	9.18	12.51	7.84	12.24
10	8.76	4.8	9.17	12.52	8.28	14.57
10	9.75	5.1	8.07	10.13	8.18	13.09
10	8.85	6.2	8.12	11.52	8.25	12.52
10	9.23	5.4	9.18	12.27	7.25	12.03
10	8.43	5.9	9.03	13.51	8.27	14.05
10	9.26	6.1	8.42	12.52	8.75	14.57
10	8.25	3.2	9.12	10.13	7.67	13.11
Average weight loss (%)	05.25		11.88		13.27	
Water absorption (%)	03.43		03.16		02.79	

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

represents a strategic approach to improve the properties and overall performance of RAC, with a particular focus on enhancing compressive strength. The treated recycled aggregates, termed STRCA, were then incorporated into concrete preparations. This innovative surface modification technique contributes to closing the existing gap in addressing durability challenges in RAC. In the abrasion treatment, a Los Angeles abrasion machine was employed, rotating at 25 revolutions per minute for 5 min to dislodge the attached mortar, following Panghal and Kumar's methodology.[64] The outcomes of the physical performance of RCA treated by abrasion are documented in Table 2.

This apparatus streamlines the process of efficiently removing mortar from the surface of the RCA. As the drum undergoes rotation, the individual aggregate particles interact with each other, initiating mechanical actions that effectively eliminate any adhered mortar. This abrasion mechanism, wherein particles rub against each other, guarantees a comprehensive cleaning of the RCA surface, rendering it free from excess mortar. Figure 2 illustrates the treatment applied to RCAs.

Subsequently, the weakened mortar adhering to the RCA surface underwent a cement slurry treatment, involving immersion in a formulated cement paste followed by drying, which aligns with Panghal and Kumar's earlier methodology.[65] To prepare the cement paste, cement is combined with water at a 10% cement-to-water ratio, ensuring an effective coating of aggregate surfaces by the paste. The mixture undergoes thorough agitation for 10–15 min to achieve homogeneity. Once the cement paste reaches readiness, abrasion-treated RCA is immersed in the slurry for 24 h, allowing the paste to be fully absorbed by the porous surfaces of the RCA. Subsequently, the coated aggregates are dried in an oven to optimize the penetration of cement particles into the aggregate structure. This controlled drying process eliminates excess moisture from the aggregates, leaving behind cement particles that have effectively permeated the porous surface. Through this meticulous combination of abrasion and cement slurry treatments, the RCA undergoes a transformative process, enhancing the mechanical strength of the aggregates and establishing a robust interfacial connection between the surrounding concrete matrix and the treated aggregates.

2.3 | Mix proportioning

To assess the efficacy of surface-treated Recycled Coarse Aggregates (RCA), six concrete mixtures (Table 3) were meticulously prepared with target strength of 25MPa.[73] The first reference mixture is (RC1) produced by utilizing natural aggregates, whereas second reference mixture is produced with substituted NCA with untreated RCA (RC 2), whereas the experimental mixtures (STRCA 25 to STRCA 100) substituted NCA with surface-treated RCA. Accurate weight batching, following IS 10262-2019,[74] ensured a consistent water-to-cement ratio of 0.50 for every combination. This methodical approach ensures uniformity in experimental conditions, allowing for a comprehensive investigation into the influence of STRCA on the characteristics of the final concrete mixes (Table 3).

2.4 | Testing methodology

A thorough examination of structural concrete encompassed a diverse array of tests (depicted in Figure 3), incorporating the analysis of microstructural features through x-ray diffraction (XRD), scanning electron microscopy (SEM), and energy-dispersive x-ray (EDX). Mechanical properties, including compressive strength, were assessed by averaging results from three specimens per mix. Long-term performance evaluations involved tests for drying shrinkage strain, electrical resistivity, and rapid chloride ion penetration.

2.4.1 | Material characterization

Prior to incorporating a diverse array of materials into the concrete mix, a crucial and meticulous step involves evaluating their compatibility through comprehensive microstructural analyses. This examination entails an in-depth scrutiny of the chemical elements, mineralogical properties, and microstructure of the materials. Employing EDX, XRD, and SEM analyses enables a discerning analysis of the constitution and structure of the materials, ensuring their suitability for concrete applications.

x-Ray diffraction

XRD is utilized for analyzing phase features and conducting mineralogical evaluations of crystalline minerals in various materials. Samples are scanned across an angular range of 3°–70° at a speed of 2 degrees per minute, with a sampling interval of 0.005 degrees at a 2° angle, using the Bruker D-8 advanced diffractometer system. Using Jade 7 XRD software, the study visualizes intensity on the y-axis during scans across the x-axis range from 0° to 70° by visually depicting peak intensities versus $2\theta°$.

Scanning electron microscopy analysis

Different materials used in the study undergo detailed characterization through SEM analysis. This involves a meticulous analysis of the surface morphology and microstructure of material samples from various

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Proportion of the concrete mixture.

S. no.	Mixture ID	NFA (kg/m^3)	NCA (kg/m^3)	RCA (kg/m^3)	STRCA (kg/m^3)	Cement (kg/m^3)	*W*/*C* ratio	Admixture (%)	Slump (mm)
1	RC 1	434.48	1501	0	0	390	0.5	4	114
2	RC 2	434.48	0	1501	0	390	05	4	70
3	STRCA 25	434.48	375.25	0	1125.75	390	0.5	4	104
4	STRCA 50	434.48	750.5	0	750.5	390	0.5	4	98
5	STRCA 75	434.48	1125.75	0	375.25	390	0.5	4	86
6	STRCA 100	434.48	0	0	1501	390	0.5	4	73

Abbreviations: NCA, natural coarse aggregates, NFA, natural fine aggregates; RCA, recycled coarse aggregate; STRCA, surface-treated recycled coarse aggregates.

combinations, conducted at 30 kV using the JSM 6610V scanning electron microscope at the University Science Instrumentation Centre (USIC) in Delhi. This rigorous analysis provides a comprehensive understanding of the intricate details of various materials, offering insightful knowledge on the overall composition and structure of materials.

Energy-dispersive x-ray analysis

The chemical composition of different materials is subjected to a complex statistical examination through EDX analysis, conducted using the SEM integrated with the Amatek EDS detector. Materials underwent detailed characterization at 30 kV using the JSM 6610 V scanning electron microscope at the USIC in Delhi. Because SEM utilizes vacuum conditions and an electron beam to form an image, special sample preparations are required. Samples are mounted on carbon tape, with a maximum size of 1 × 1 cm. All water must be removed to prevent vaporization in the vacuum. Electrically conductive samples require no further preparation, but non-conductive samples must be made conductive by coating them with a thin layer of Au using a sputter coater. This application adds a level of accuracy to the characterization process by analyzing the presence of distinct constituents in different materials using both qualitative and quantitative evaluations. Collectively, these analyses contribute to a comprehensive assessment of material suitability, ensuring the effective integration of components in the various materials.

2.4.2 | Compressive strength

Thirty meticulously crafted specimens, contained within steel cube molds measuring 150 × 150 × 150 mm, underwent meticulous preparation, curing, and subsequent compressive strength testing. Utilizing compression testing equipment with a capacity of 2000 kN, the average compressive strength values for each combination were calculated based on data from three specimens. These comprehensive assessments conducted over both 7 and 28 days offer a profound understanding of the strength properties exhibited by diverse concrete mixes at different points in time.

2.4.3 | Drying shrinkage

The evaluation of drying shrinkage in concrete mixes adhered to a systematic observation of length changes over time, following the IS 516 (Part 6) 2020[75] standards. For each of the five mixtures, three specimens measuring 75 × 75 × 285 mm were stored for a duration of 28 days. Initial measurements were meticulously recorded with a precision of 0.005 mm. Subsequent readings at 28, 56, and 90 days were obtained using dial gages. The alteration in length for each specimen was calculated by assessing the disparity between the final and initial values, providing a comprehensive understanding of the drying shrinkage characteristics over an extended duration.

2.4.4 | Electrical resistivity

The assessment of concrete electrical resistivity plays a vital role in comprehending potential corrosion risks to embedded steel reinforcement bars. This method entails measuring the surface resistivity of fresh concrete using a four-electrode setup and a Resipod Resistivity Meter. The samples meet RILEM TC 154-EMC[76] requirements and are made in 100 mm diameter by 200 mm height cylindrical molds. Evaluations conducted at 28 and 56 days

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



FIGURE 3 Evaluations of (a) compressive strength, (b) drying shrinkage, (c) electrical resistivity, and (d) rapid chloride penetration test setup.

provide comprehensive insights into the evolution of concrete electrical resistivity over time, contributing to a thorough analysis of corrosion susceptibility.

2.4.5 | Rapid chloride penetration

Conforming to the American Society for Testing and Materials (ASTM) C1202[77] standards, the mold of the rapid chloride penetration test (RCPT) apparatus has a 100 mm diameter and a 50 mm height. Thirty specimens were meticulously prepared for this study. The unit incorporates two sealed chambers containing sodium chloride (NaCl) at 0.24 M and sodium hydroxide (NaOH) at 0.3 M. A 60 V current source connects to the NaCl and NaOH cells, with readings recorded at 30-min intervals for up to 6 h. Tests conducted at 28 and 56 days measure the charge passed, providing a crucial indicator of concrete's resistance to chloride ion penetration.

3 | RESULTS AND DISCUSSION

3.1 | Materials characterization

The examination of material microstructure, such as cement and aggregates, with concrete involves a crucial preliminary step of microstructural analysis before

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

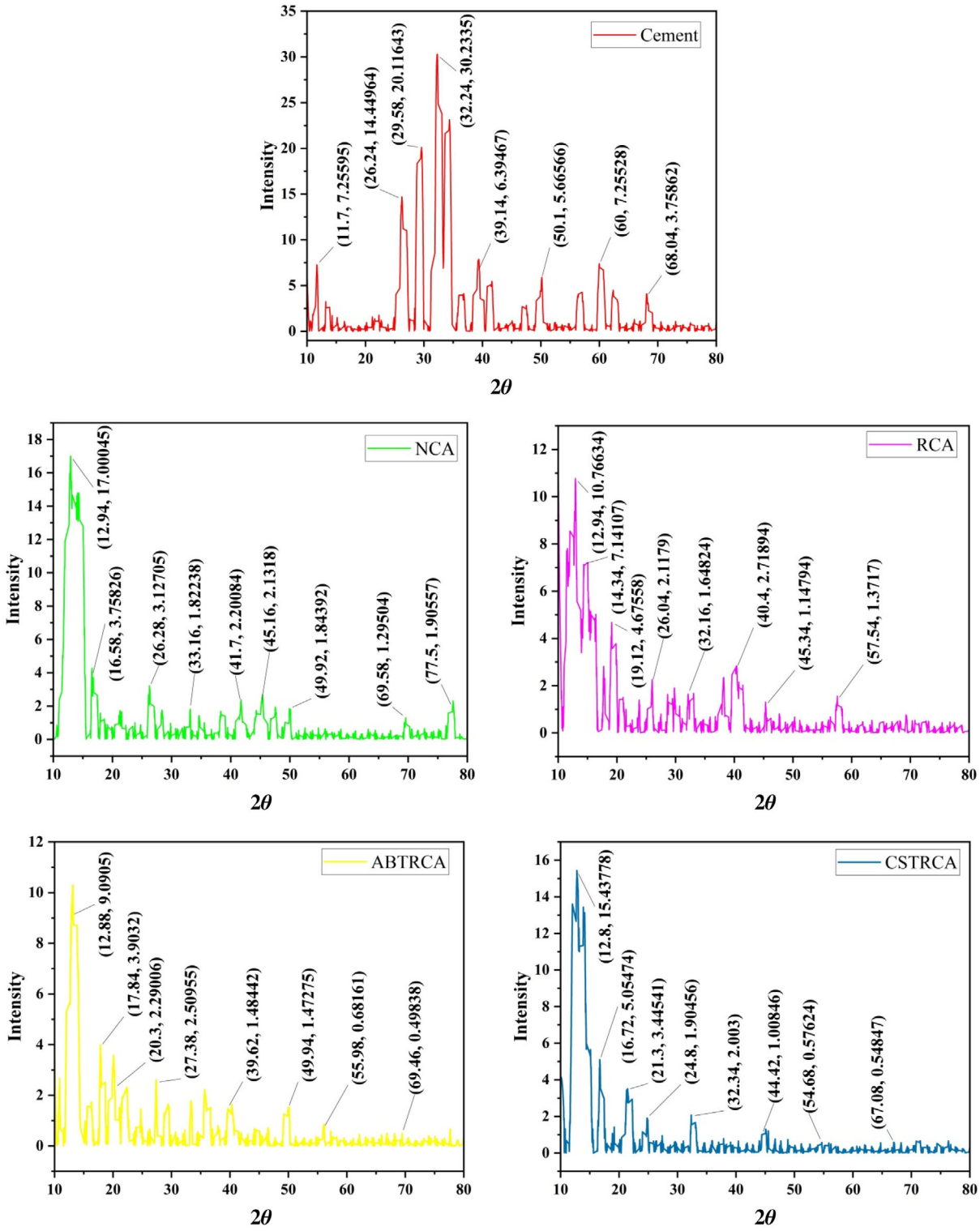


FIGURE 4 x-Ray diffraction data comparison of cement, natural coarse aggregates (NCA), recycled coarse aggregate (RCA), abrasion-treated recycled coarse aggregates (ABTRCA), and cement slurry-treated recycled coarse aggregates (CSTRCA).

incorporation. To comprehensively evaluate the chemical elements, mineralogical properties, and microstructure of these materials, an in-depth analysis was conducted using XRD, SEM, and EDX. These analytical methods provide thorough insights into the structure and composition of the materials, establishing a foundational understanding of their suitability in concrete applications. The ensuing discussion will delve into the implications of these findings on the overall performance and durability of the concrete mixtures.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 x-Ray diffraction data for various materials.

S. no.	Material	Peak no.	Peak angle	Intensity	Chemical formula	Chemical name	Crystal system
1	Cement	1	32.24	30.2335	C_3S	Tri calcium silicate	Monoclinic and rhombohedra
		2	29.58	20.11643	C_2S	Di calcium silicate	Octahedral and hexahedron
		3	26.24	14.44964	C_3A	Tri calcium aluminate	Cubic and monoclinic
		4	11.7	7.25595	C_4AF	Tetra calcium aluminate ferrate	Octahedral and tetrahedral
		5	39.14	6.39467	CaO	Calcium oxide	Cubic
		6	50.1	5.66566	MgO	Periclase	Octahedral
		7	68.04	3.75862	$CaSO_4$	Calcium sulfate	Orthorhombic
2	NCA	1	12.94	17.00045	SiO_2	Quartz	Tetrahedral
		2	16.58	3.75826	Hbl	Hornblende	Double tetrahedral
		3	26.28	3.12705	$NaAlSi_3O_8$	Albite	Tetrahedral
		4	33.16	1.82238	$CaAl_2Si_2O_8$	Anorthite	Triclinic with prismatic
		5	41.7	2.20084	H-M	Chamosite	Prismatic
		6	45.16	1.29504	Fe_2O_3	Hematite	Hexagonal
		7	77.5	1.90557	$CaMg(CO_3)_2$	Dolomite	Trigonal
3	RCA	1	12.94	10.76634	SiO_2	Quartz	Tetrahedral
		2	14.34	7.14107	Hbl	Hornblende	Double tetrahedral
		3	19.12	4.67558	$NaAlSi_3O_8$	Albite	Tetrahedral
		4	26.04	2.1179	$CaAl_2Si_2O_8$	Anorthite	Triclinic with prismatic
		5	32.16	1.6482	H-M	Chamosite	Prismatic
		6	45.34	1.14794	Fe_2O_3	Hematite	Hexagonal
		7	57.54	1.3717	$CaMg(CO_3)_2$	Dolomite	Trigonal
4	ABTRCA	1	12.88	9.0905	SiO_2	Quartz	Tetrahedral
		2	17.84	3.9032	Hbl	Hornblende	Double tetrahedral
		3	27.38	2.50955	$NaAlSi_3O_8$	Albite	Tetrahedral
		4	39.62	1.48442	$CaAl_2Si_2O_8$	Anorthite	Triclinic with prismatic
		5	49.94	1.47275	H-M	Chamosite	Prismatic
		6	55.98	0.68161	Fe_2O_3	Hematite	Hexagonal
		7	69.46	0.49838	$CaMg(CO_3)_2$	Dolomite	Trigonal
5	CSTRCA	1	12.8	15.43778	SiO_2	Quartz	Tetrahedral
		2	16.72	5.05474	Hbl	Hornblende	Double tetrahedral
		3	21.3	3.44541	$NaAlSi_3O_8$	Albite	Tetrahedral
		4	24.8	1.90456	$CaAl_2Si_2O_8$	Anorthite	Triclinic with prismatic
		5	32.34	2.003	H-M	Chamosite	Prismatic
		6	44.42	1.00846	Fe_2O_3	Hematite	Hexagonal
		7	67.08	0.54847	$CaMg(CO_3)_2$	Dolomite	Trigonal

Abbreviations: ABTRCA, abrasion-treated recycled coarse aggregates; CSTRCA, cement slurry-treated recycled coarse aggregates; NCA, natural coarse aggregates, RCA, recycled coarse aggregate.

3.1.1 | XRD analysis of materials

In Figure 4, the XRD results for cement reveal distinct peaks corresponding to calcium silicate, tricalcium aluminate, and silica phases. Table 4 provides comprehensive details regarding XRD peaks, including their positions, intensities, chemical nomenclature, chemical formulas, and crystal systems for all materials. Moving

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

on to NCA in Figure 4, the XRD analysis focuses on selected chemical elements, including calcium (Ca), oxygen (O), silicon (Si), aluminum (Al), iron (Fe), potassium (K), sodium (Na), and magnesium (Mg). The subsequent XRD analysis of RCA (Figure 4) unveils its mineralogical properties, showcasing peaks corresponding to quartz, hornblende, albite, and anorthite. Notably, the nature of RCA appears crystalline prior to surface modification treatment, indicating that the aggregate surface has ancient cement mortar adhered to it, potentially leading to weak bonding with the cement matrix. The unmodified RCA particles exhibit insufficient bonding, prompting the need for their removal to establish robust bonding with the cement matrix. To address this, abrasion and cement slurry treatments are applied for surface modification. The XRD analysis of STRCA, following abrasion and cement slurry treatments, presents mineralogical properties in Figure 4.

Notably, a significant peak indicates the prevalence of quartz in the majority of the treated material. Abrasion and cement slurry treatments induce observable changes in the XRD patterns of RCAs. These changes include discernible alterations in peak intensities, indicating the dissolution of minerals present on the aggregates' surface. The emergence of new peaks suggests the potential formation of novel compounds previously obscured by impurities. Amorphization, reflected in broader peaks, indicates a shift from crystalline to amorphous structures. Crystal structure modifications and shifts in peak positions signify alterations in specific mineral phases. The abrasion treatment may also result in mineral depletion, removing some minerals entirely from the surface. These variations collectively enhance our understanding of how abrasion treatment influences the mineralogical composition of RCAs, as discerned through XRD analysis. These results highlight how well the surface treatments work to improve the mineralogical features of RCA, thereby contributing to improved bonding with the cement matrix. A comparable observation is also noted by Kazemian et al.,[59] as indicated by the reported findings. The ensuing discussion will delve into the implications of these mineralogical variations for the overall performance and durability of the concrete mixtures. The STRCA exhibits a rougher surface, closely resembling the characteristics of NCA. This distinction suggests that STRCA is more conducive to forming a robust bond with the cement matrix, thereby enhancing the mechanical characteristics of RAC.

3.1.2 | SEM analysis of materials

SEM was employed for the characterization of various materials, including cement, NCA, RCA, abrasion-treated recycled coarse aggregates (ABTRCA), cement slurry-treated recycled coarse aggregates (CSTRCA). This analytical instrument allows for the examination of the surface morphology and particle microstructure of these materials. As depicted in Figure 5, SEM images provide insights into the surfaces of NCA, RCA, and STRCA samples. The SEM image of the NCA sample reveals a dense, uniform surface, relatively free of impurities except for minor dust resulting from the manufacturing crushing process. In contrast, the unmodified RCA sample exhibits a more porous surface covered with fine particles and loose cement mortar, a consequence of the crushing process, as observed in the figure. A similar observation is reported by Ismail and Ramli.[30,58] SEM images of the surface of CSTRCA samples depict the covering with cement particles. The abrasion treatment effectively attacked the RCA surface, dissolving the adhered mortar and creating new pores, as seen in SEM images of ABTRCA. These surface modifications, attributed to the cement slurry treatment, contribute to increased bonding with cement paste and enhance the overall strength of the mixture. ABTRCA and CSTRCA exhibit a cleaner, denser, and more uniform surface compared to the unmodified RCA, as visually represented in Figure 7.

3.1.3 | EDX analysis of materials

Thorough analysis using EDX was conducted to identify the crystalline constituents within both cement powder and coarse aggregates, following the stringent guidelines outlined by ASTM. Figure 6 presents the outcomes of the EDX analysis for various materials, including cement, NCA, RCA, ABTRCA, and CSTRCA. The comprehensive EDX analysis distinctly illustrates the presence of all chemical elements in each material, forming a foundational understanding of their elemental composition. The ensuing discussion will delve into the implications of these findings for the structural and chemical properties of the concrete mixtures. The figures clearly illustrate a noticeable rise in Si percentages following surface modification. This uptick signifies the enhanced distribution of calcium silicate minerals on the surface of RCAs, characterized by the presence of Si, O, and Ca. The higher percentage of this distribution has a major role in the enhanced cement paste-aggregate bond. Thus, the material's mechanical characteristics have increased, and this reinforced link is to blame. Kessal et al.[51] also make a comparable observation, according to the reported findings.

3.2 | Compressive strength

The results presented in Table 5 and Figure 7 elucidate the percentage variation in compressive strength among

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

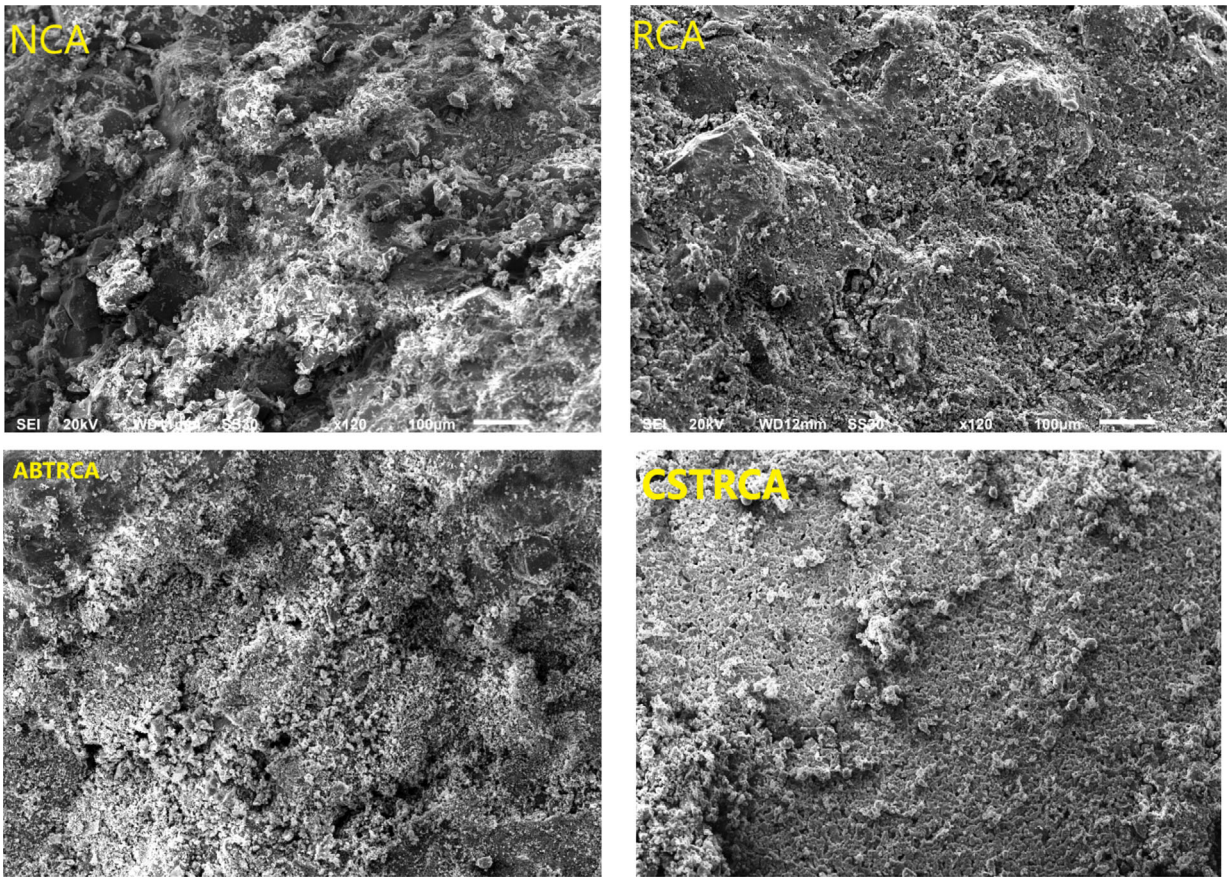


FIGURE 5 Scanning electron microscopy analysis results for natural coarse aggregates (NCA), recycled coarse aggregate (RCA), abrasion-treated recycled coarse aggregates (ABTRCA), and cement slurry-treated recycled coarse aggregates (CSTRCA).

various concrete mixtures at 7 and 28 days, referencing the control mix (RC 1), all characterized by medium workability (50–100 mm). Concrete mixes incorporating increasing amounts of STRCA exhibit lower compressive strength than the reference mix RC 1, and strength levels ascend with the curing age. Notably, among the evaluated surface modifications, a 50% replacement emerges as the optimal choice. This implies that substituting 25% of regular aggregate with STRCA yields the highest strength after 28 days. This enhancement stems from the utilization of recycled aggregate, effectively eliminating outdated mortar and resulting in a mix denoted as STRCA 25, which offers maximum strength with reduced voids and higher particle density. Compressive strength serves as a pivotal parameter in evaluating concrete performance, and variations in strength offer insights into the efficacy of incorporating supplementary materials. The control mix RC 1 demonstrated a compressive strength of 23.01 MPa at 7 days and 33.91 MPa at 28 days. In comparison, the control mix RC 2 demonstrated a compressive strength of 16.31 MPa at 7 days and 22.91 MPa at 28 days. Subsequent mixtures integrated STRCA at different replacement levels—25%, 50%, 75%, and 100%.

At 7 days, the STRCA 25 mixture showcased a noteworthy increase in compressive strength, reaching 29.95 MPa, representing a positive variation of 30.16% compared to the RC 1 mix. This suggests that the inclusion of 25% recycled concrete aggregate positively influenced early-age strength development. Similarly, STRCA 50 and STRCA 75 mixtures displayed positive variations of 18.99% and 1.59%, respectively, indicating continued enhancement in compressive strength compared to the control mix (RC 1) at 7 days. However, STRCA 100 exhibited a decrease in strength, registering 19.18 MPa and a negative variation of 16.64%, indicating that a complete replacement of natural aggregates with recycled material adversely impacted early-age compressive strength. As the curing period extended to 28 days, the trends in compressive strength variations persisted with some modifications. STRCA 25, STRCA 50, and STRCA 75 maintained positive variations of 29.37%, 17.13%, and 1.59%, respectively, compared to the control mix (RC 1). These results indicate that the positive influence of recycled concrete aggregate on compressive strength was sustained over the curing period. However, STRCA 100 continued to exhibit a decrease in strength with a negative variation of 11.58%, suggesting a long-term

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

adverse effect on compressive strength when natural aggregates are entirely replaced with recycled material. According to the study findings, STRCA 25 demonstrates the highest compressive strength at 43.87 MPa, while STRCA 100 exhibits the lowest at 29.98 MPa. The results in Table 5 show that all mixtures produced with integrated STRCA at various replacement levels (25%, 50%, 75%, and 100%) performed significantly better than the control mix RC 2, which was produced with untreated RCA substituted for NCA at a 100% replacement level.

To gain a thorough and comprehensive understanding of the results obtained from our compressive strength tests, it is crucial to closely analyze the combined aspects of the loading rate and the force–displacement curve. The loading rate, which refers to the speed at which the load is applied to the test specimen, significantly influences the material's response under stress. Simultaneously, the force–displacement curve provides a detailed graphical representation of how the material deforms under various loads, showcasing the relationship between the applied force and the resulting displacement. By examining this aspect, we can gain valuable insights into the material's behavior under compression, identify any potential weaknesses, and better understand its mechanical properties. Figure 8 provides a detailed illustration of these parameters, offering a visual aid to support the analysis and interpretation of the test results. This figure is essential for grasping the nuances of the material's performance and for making informed decisions based on the test outcomes. These results align with those reported

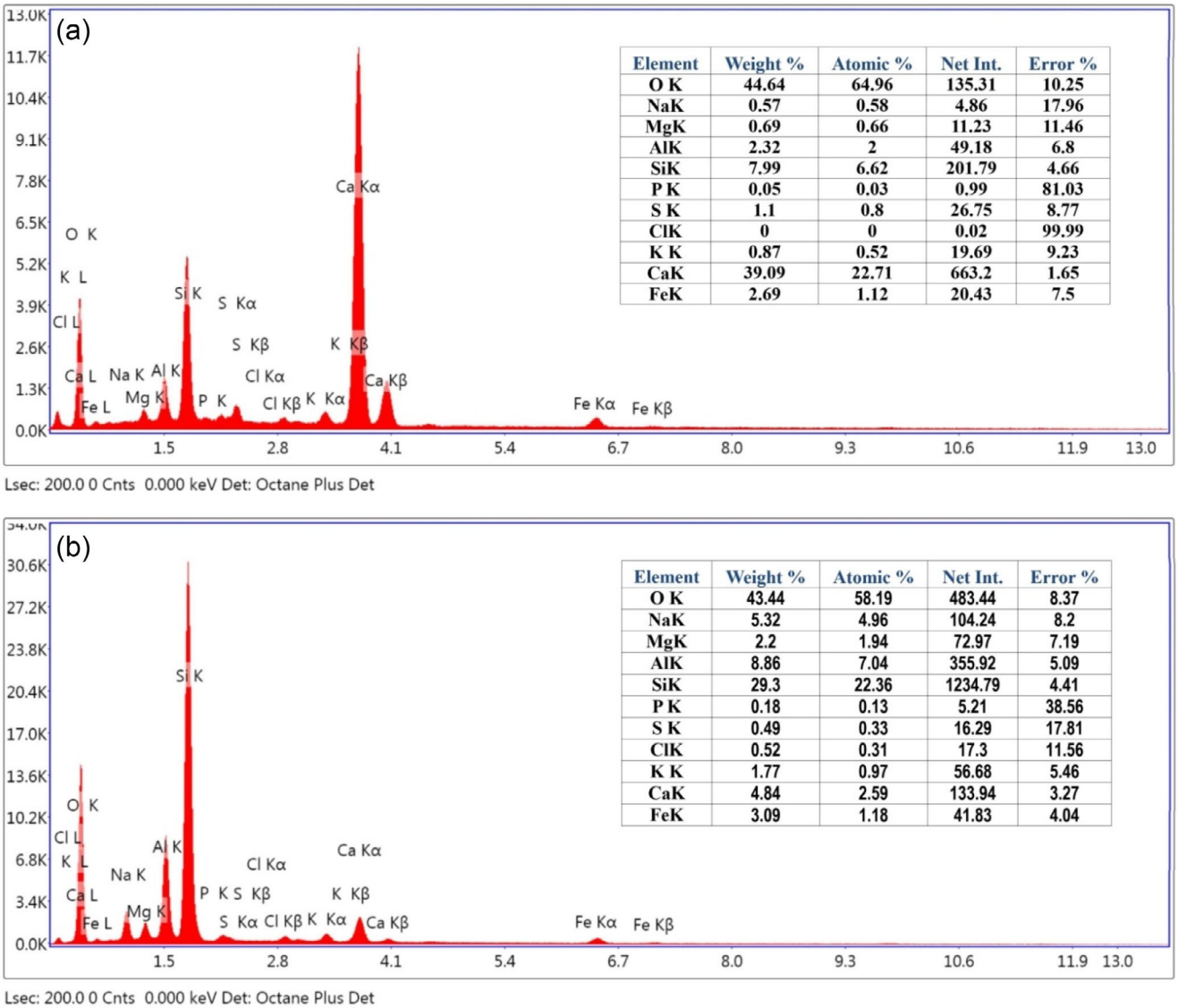
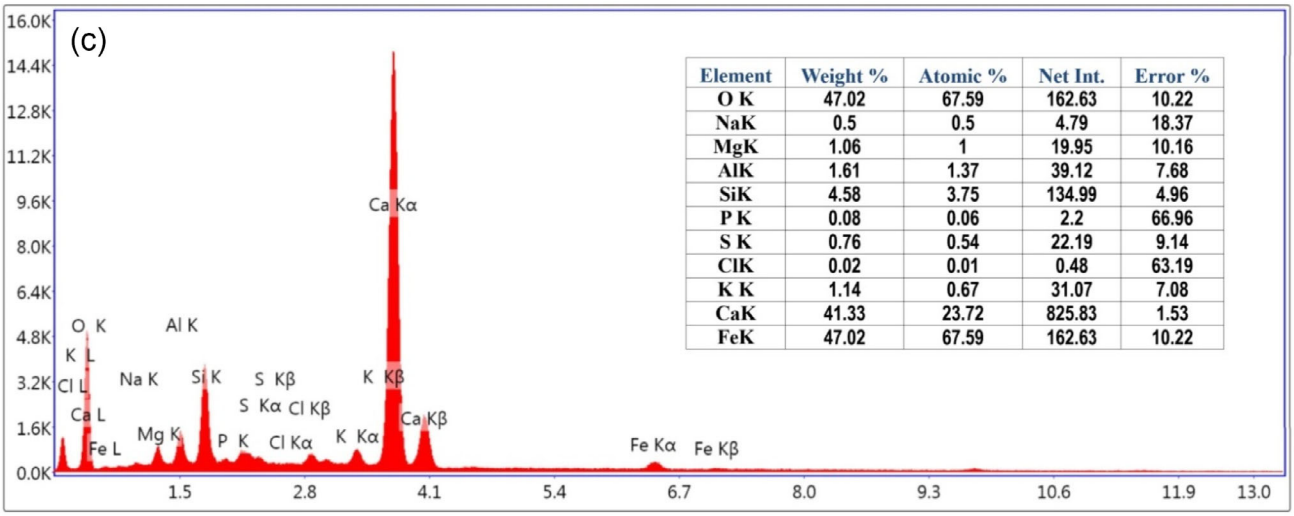
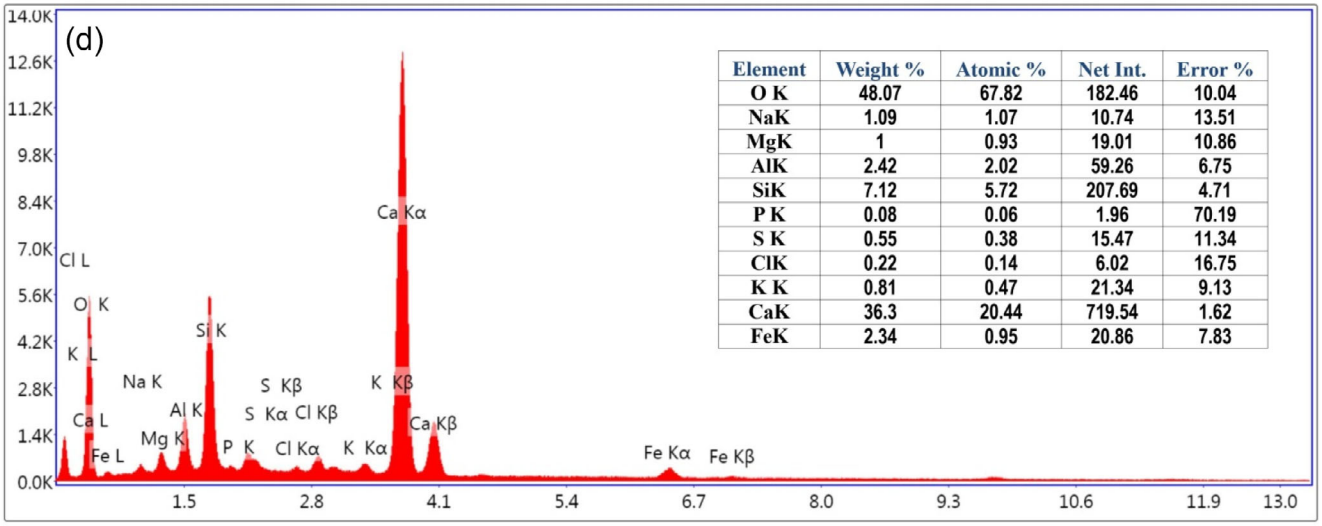


FIGURE 6 Energy-dispersive x-ray analyses for (a) cement, (b) natural coarse aggregates, (c) recycled coarse aggregate, (d) abrasion-treated recycled coarse aggregates, and (e) cement slurry-treated recycled coarse aggregates.

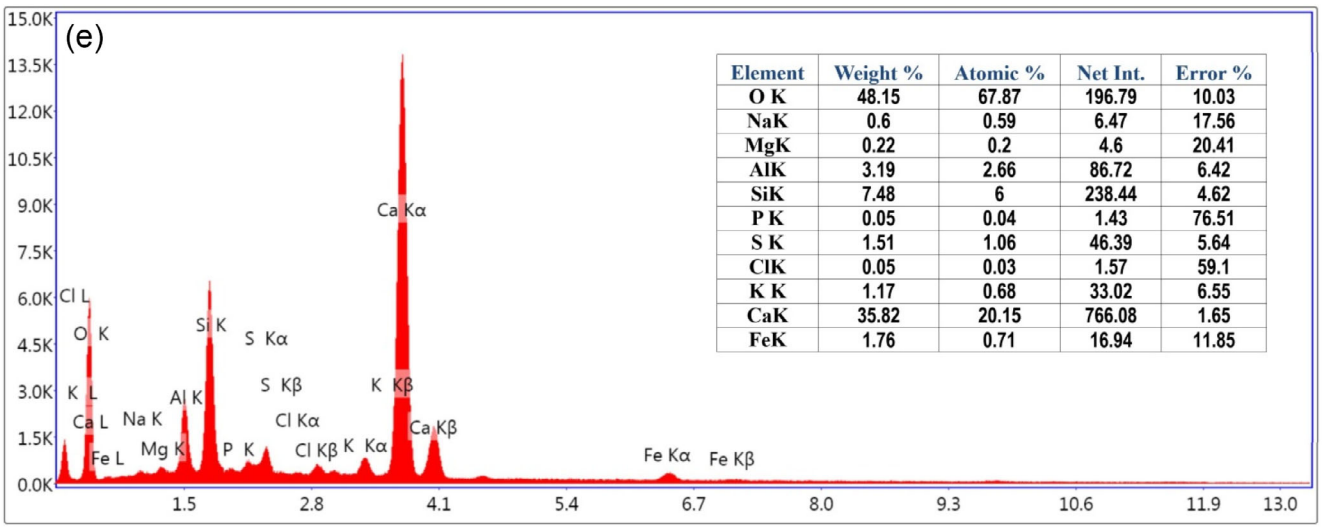
17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Element	Weight %	Atomic %	Net Int.	Error %
O K	47.02	67.59	162.63	10.22
NaK	0.5	0.5	4.79	18.37
MgK	1.06	1	19.95	10.16
AlK	1.61	1.37	39.12	7.68
SiK	4.58	3.75	134.99	4.96
P K	0.08	0.06	2.2	66.96
S K	0.76	0.54	22.19	9.14
ClK	0.02	0.01	0.48	63.19
K K	1.14	0.67	31.07	7.08
CaK	41.33	23.72	825.83	1.53
FeK	47.02	67.59	162.63	10.22



Element	Weight %	Atomic %	Net Int.	Error %
O K	48.07	67.82	182.46	10.04
NaK	1.09	1.07	10.74	13.51
MgK	1	0.93	19.01	10.86
AlK	2.42	2.02	59.26	6.75
SiK	7.12	5.72	207.69	4.71
P K	0.08	0.06	1.96	70.19
S K	0.55	0.38	15.47	11.34
ClK	0.22	0.14	6.02	16.75
K K	0.81	0.47	21.34	9.13
CaK	36.3	20.44	719.54	1.62
FeK	2.34	0.95	20.86	7.83



Element	Weight %	Atomic %	Net Int.	Error %
O K	48.15	67.87	196.79	10.03
NaK	0.6	0.59	6.47	17.56
MgK	0.22	0.2	4.6	20.41
AlK	3.19	2.66	86.72	6.42
SiK	7.48	6	238.44	4.62
P K	0.05	0.04	1.43	76.51
S K	1.51	1.06	46.39	5.64
ClK	0.05	0.03	1.57	59.1
K K	1.17	0.68	33.02	6.55
CaK	35.82	20.15	766.08	1.65
FeK	1.76	0.71	16.94	11.85

FIGURE 6 (Continued)

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

by Kessal et al.[51] Additionally, Wagih et al.[78] observed a notable reduction in compressive strength by 20%–34% at 7 days and 18%–28% at 28 days when using RCA in concrete mixes. However, after applying surface treatment, the decrease was limited to only 04%–17% at 7 days and 02%–12% at 28 days, even with 100% replacement of RCA. Similar positive outcomes were noted in an earlier study conducted by Wang et al.[32] and Kazemian et al.[59] In summary, the variations in compressive strength for different mixtures indicate that the inclusion of recycled concrete aggregate at lower replacement levels positively influences both early-age and long-term strength, while complete replacement may result in a decline in compressive strength. These findings bear significant implications for the sustainable use of recycled materials in concrete production, emphasizing the necessity for careful consideration of replacement levels to optimize both short-term and long-term performance.

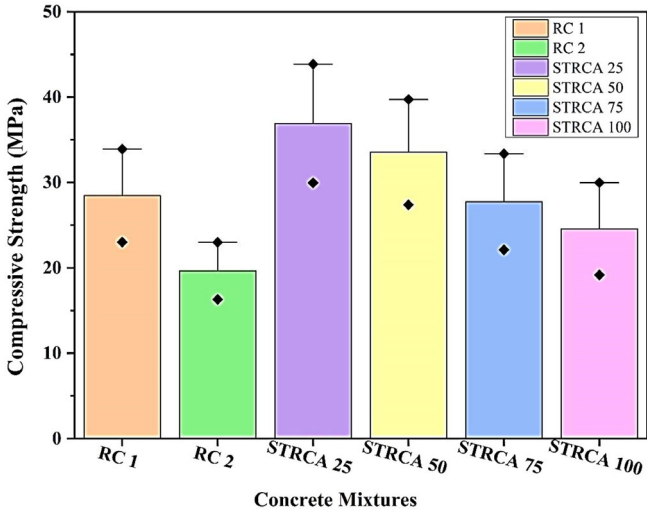


FIGURE 7 Variations in compressive strength across different concrete mixtures. STRCA, surface-treated recycled coarse aggregates.

3.3 | Drying shrinkage

Drying shrinkage assessment is crucial for understanding structural concrete performance. Table 6 and Figure 9 depict percentage variations in drying shrinkage for different concrete mixtures at 28, 56, and 90 days, referencing the control mix (RC 1). This parameter is vital for evaluating cracking and deformations. Results show that all mixtures align with Indian standards, implying that incorporating STRCA is viable within defined limits. The drying shrinkage curve transitions from steep to flat with prolonged drying time, indicating a diminishing rate of change. Higher STRCA replacement increases drying shrinkage, but it also positively influences and mitigates it compared to untreated RCA. This emphasizes the strategic use of STRCA to balance sustainability without compromising concrete performance. The control mix RC 1 exhibited drying shrinkage values of 348×10^{-6} at 28 days, 431×10^{-6} at 56 days, and 474×10^{-6} at 90 days. In contrast, the control mix RC 2 exhibited

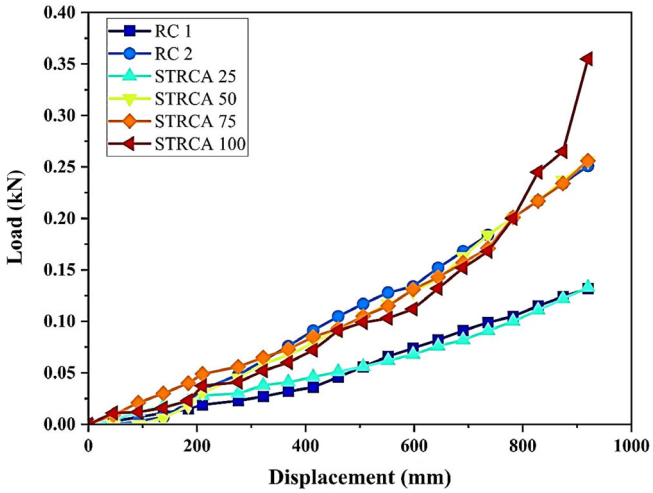


FIGURE 8 Load–displacement curve for different mixtures.

TABLE 5 Percentage variation of compressive strength for different mixtures.

	Compressive strength (MPa)					
Concrete mix	7 Days	% variation to RC 1	% Variation to RC 2	28 Days	% Variation to RC 1	% Variation to RC 2
RC 1	23.01	-	-	33.91	-	-
RC 2	16.31	−29.11	-	22.99	−32.20	-
STRCA 25	29.95	+30.16	+83.62	43.87	+29.37	+90.82
STRCA 50	27.38	+18.99	+67.87	39.72	+17.13	+72.77
STRCA 75	22.11	−03.91	+35.56	33.37	+01.59	+45.15
STRCA 100	19.18	−16.64	+17.60	29.98	−11.58	+30.40

Note: + Sign represents an increase in strength and – sign represents a decrease in strength.
Abbreviation: STRCA, surface-treated recycled coarse aggregates.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

drying shrinkage values of 499×10^{-6} at 28 days, 624×10^{-6} at 56 days, and 641×10^{-6} at 90 days. Subsequent mixtures incorporated STRCA at various replacement levels (25%, 50%, 75%, and 100%). At 28 days, STRCA 25 showed a slight increase in drying shrinkage (354×10^{-6}), representing a positive variation of 01.72% compared to the RC 1 mix, indicating minimal impact at this early stage. However, as the percentage of STRCA increased, drying shrinkage also rose, with positive variations of 10.91%, 25.86%, and 38.79% for STRCA 50, STRCA 75, and STRCA 100, respectively, at 28 days.

At 56 days, STRCA 25, STRCA 50, STRCA 75, and STRCA 100 exhibited percentage variations of +02.32%, +10.90%, +26.68%, and +40.60%, respectively, at 56 days. These results indicate a sustained influence of STRCA on increasing drying shrinkage over time, with higher replacement levels contributing to more significant variations. Moving to 90 days, STRCA 25, STRCA 50, STRCA 75, and STRCA 100 showed percentage variations of +02.32%, +10.97%, +22.15%, and +31.85%, respectively, at 90 days. The results in Table 6 show that all mixtures produced with integrated STRCA at various replacement levels (25%, 50%, 75%, and 100%) have lower drying shrinkage values than the control mix RC 2, which was produced with untreated RCA substituted for NCA at a 100% replacement level. The positive variations observed in drying shrinkage imply that the inclusion of STRCA, especially at higher replacement levels, contributes to an increase in the potential for cracking and deformations in the concrete. This information is crucial for structural engineers and concrete designers as they consider the long-term durability and performance of concrete structures. Furthermore, STRCA proves advantageous in minimizing drying shrinkage compared to unmodified RCA, attributed to the efficacy of the surface modification process applied to RCAs. This process substantially diminishes mortar presence on the aggregate surface, a pivotal factor in augmented water absorption and ensuing drying shrinkage. Consequently, the removal of attached mortar through surface modification results in significantly reduced drying shrinkage compared to unmodified RCA, underscoring the crucial role of surface modification in enhancing the performance of STRCA.

The study concludes that at 100% replacement, STRCA exhibits an 11%–20% reduction in shrinkage strain compared to unmodified RCA. This finding aligns with research conducted by Ismail and Ramli,[30] exploring how treated recycled aggregates influence the drying shrinkage of concrete. Prior research by Kioumarsi et al.[79] indicated a higher drying shrinkage strain with unmodified RCAs, consistent with observations by Wang et al.,[32] suggesting a systematic increase in shrinkage with a higher proportion of unmodified aggregates. In summary, the results presented in Figure 8 indicate that the drying shrinkage of concrete increases with higher replacement levels of STRCA. While lower levels (STRCA 25) show minimal impact, significant positive variations in drying shrinkage are observed at higher replacement levels (STRCA 50, STRCA 75, and STRCA 100). These findings underscore the importance of carefully balancing the use of recycled materials in concrete mixtures to mitigate the risk of increased drying shrinkage and its potential consequences on the long-term performance of concrete structures.

3.4 | Electrical resistivity

The connectivity of the porous network within cement mortar directly influences electrical conductivity,

TABLE 6 Percentage variations in drying shrinkage among different mixtures.

	Drying shrinkage (10^{-6})								
Concrete mix	28 Days	% Variation with respect to RC 1	% Variation with respect to RC 2	56 Days	% Variation with respect to RC 1	% Variation with respect to RC 2	90 Days	% Variation with respect to RC 1	% Variation with respect to RC 2
RC 1	348	-	-	431	-	-	474	-	-
RC 2	499	+53.54	-	624	+55.61	-	641	+45.68	-
STRCA 25	354	+01.72	−29.05	441	+02.32	−29.32	485	+02.32	−24.33
STRCA 50	386	+10.91	−22.64	478	+10.90	−24.67	526	+10.97	−17.94
STRCA 75	438	+25.86	−12.22	546	+26.68	−12.50	579	+22.15	−09.67
STRCA 100	483	+38.79	−03.20	606	+40.60	−02.88	625	+31.85	−02.49

Note: + Sign represents increase in drying shrinkage and – sign represents decrease in drying shrinkage.
Abbreviation: STRCA, surface-treated recycled coarse aggregates.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

providing a precise measure of ionic transport resistance. Consequently, electrical resistivity serves as a crucial metric for assessing concrete quality. Table 7 and Figure 10 illustrate the percentage variations in electrical resistivity for diverse concrete mixtures at 28 and 56 days, using the control mix (RC 1) as a reference. The outcomes reveal a substantial improvement in the electrical resistance of cement mortar following surface modification. In this study, the primary factor contributing to decreased electrical conductivity appears to be the integrated surface modification approach applied to the cement mortar. This modification is presumed to play a pivotal role in altering the pore network, enhancing electrical resistance. At 28 days, RC 1 exhibited an electrical resistivity of 47.45 kΩ. Mixtures with STRCA at varying replacement levels showed diverse effects on electrical resistivity. STRCA 25 displayed a positive variation of +03.41%, indicating a slight increase, while STRCA 50 showed a marginal rise of +01.26%. However, STRCA 75 and STRCA 100 exhibited negative variations of −15.44% and −26.25%, suggesting decreased electrical resistivity. As the curing period extended to 56 days, the trends in electrical resistivity variations continued. The control mix, RC 1, had an electrical resistivity of 58.91 kΩ. STRCA 25 and STRCA 50 displayed positive variations of +00.88% and +01.74%, respectively, indicating slight increases compared to the control mix. However, STRCA 75 and STRCA 100 showed negative variations of −12.47% and −22.23%, respectively, signifying a notable decrease in electrical resistivity. The control mix RC 2 exhibited electrical resistivity values of 16.01 kΩ at 28 days and 26.29 kΩ at 56 days. The results in Table 7 show that all mixtures produced with integrated STRCA at various replacement levels (25%, 50%, 75%, and 100%) have higher electrical resistivity values

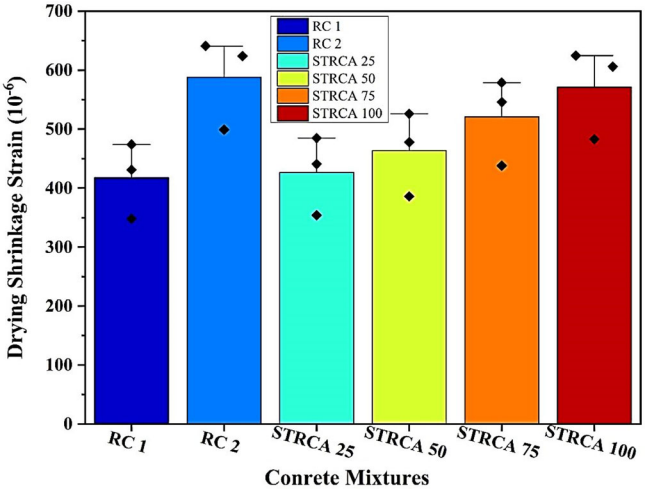


FIGURE 9 Shrinkage strain comparison among different concrete mixtures. STRCA, surface-treated recycled coarse aggregates.

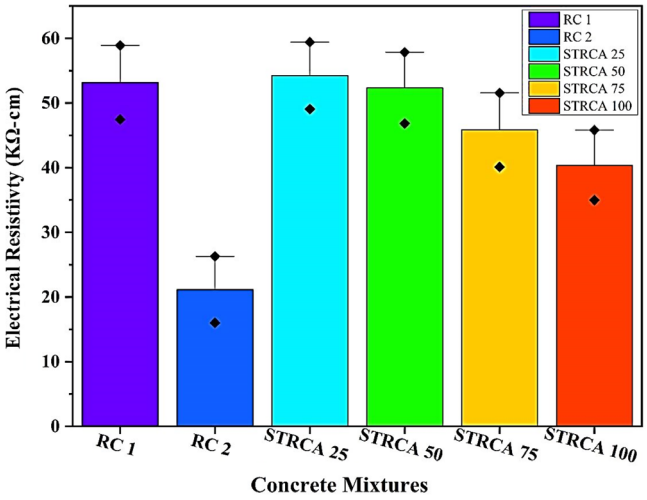


FIGURE 10 Fluctuations in electrical resistivity across various concrete mixtures. STRCA, surface-treated recycled coarse aggregates.

TABLE 7 Percentage variations in electrical resistivity for different mixtures.

Concrete mix	Electrical resistivity (kΩ)					
	28 Days	% Variation with respect to RC 1	% Variation with respect to RC 2	56 Days	% Variation with respect to RC 1	% Variation with respect to RC 2
RC 1	47.45	-	-	58.91	-	-
RC 2	16.01	−29.50	-	26.29	−17.82	-
STRCA 25	49.07	+03.41	+206.49	59.43	+00.88	+126.05
STRCA 50	46.85	+01.26	+192.62	57.88	+01.74	+120.15
STRCA 75	40.12	−15.44	+150.59	51.56	−12.47	+96.120
STRCA 100	34.99	−26.25	+118.55	45.81	−22.23	+74.24

Note: + Sign represents increase in electrical resistivity and – sign represents decrease in electrical resistivity.
Abbreviation: STRCA, surface-treated recycled coarse aggregates.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

than the control mix RC 2, which was produced with untreated RCA substituted for NCA at a 100% replacement level.

In summary, STRCA inclusion had varied effects on electrical resistivity, with lower replacements showing marginal increases or maintenance, while higher replacements led to significant decreases. The results underscore the need to carefully evaluate electrical properties when incorporating recycled materials, which are crucial for concrete durability and corrosion resistance. In the context of these findings, the incorporation of STRCA significantly influences the electrical resistivity of cement mortar. The observed variations suggest that STRCA surface modification effectively alters the concrete's porous network, contributing to enhanced resistance to ionic transport. The study concludes that at 100% replacement, STRCA exhibits a 28%–36% increment in electrical resistance compared to unmodified RCA. These findings align with studies by Sasanipour and Aslani,[52] investigating the electrical resistivity of concrete with pre-treated RCA. Similarly, research by Zhan et al.[36] exploring treated recycled aggregates aligns with our study, suggesting that concrete with STRCA poses a low risk of corrosion to steel bars. These results highlight the potential of STRCA as a beneficial component in concrete mixtures, particularly for applications requiring improved electrical resistivity. Consequently, concrete with up to 100% RCA replacement carries a low risk of corrosion.

3.5 | Rapid chloride penetration

The evaluation of concrete's chloride resistance relies on the total charge measured by the RCPT, offering insights into chloride ion transport characteristics. Table 8 and Figure 11 illustrate the percentage variations in passed charge for diverse concrete mixtures at 28 and 56 days, using the control mix (RC 1) as a baseline. Passed charge is a crucial parameter in appraising the electrical performance of concrete, providing indications of its ability to conduct electrical current. Remarkably, the outcomes suggest that surface modification positively influences chloride ion penetration resistance. The inclusion of STRCA contributes to establishing a robust bond between aggregates and cement paste, fostering the formation of strong interfacial transition zones (ITZs). These enhanced ITZs play a pivotal role in impeding chloride ion penetration, owing to the cleaner, denser, and more uniformly shaped STRCA acting as a barrier against chloride ion ingress. This signifies a substantial enhancement in concrete's ability to withstand the deleterious effects of chloride exposure, highlighting the positive influence of surface modification on chloride resistance. At 28 days, the control mix, RC 1, recorded a passed charge of 704 C. Subsequent mixtures with STRCA at various replacement levels displayed variations. STRCA 25 showed a slight decrease of −00.56%, indicating a marginal reduction. In contrast, STRCA 50 exhibited a positive variation of +04.54%, suggesting an increase. Notably, STRCA 75 and STRCA 100 demonstrated significant positive variations of +11.78% and +22.72%, respectively, indicating substantial increases in passed charge at 28 days. As the curing period extended to 56 days, the trends persisted. RC 1 had a passed charge of 683 C. STRCA 25 exhibited a marginal decrease of −00.43%, while STRCA 50 and STRCA 75 showed further positive variations of +05.85% and +16.54%, respectively. STRCA 100 displayed the highest positive variation of +28.69%, indicating a significant increase in passed charge at 56 days. The control mix RC 2 demonstrated a passed charge of 1053 C at 28 days and 1050 C at 56 days. As illustrated in Table 8, all mixtures incorporating STRCA at varying replacement levels (25%, 50%, 75%, and 100%) exhibited lower passed charges, indicating higher resistance values compared to

TABLE 8 Percentage variations in passed charge for different mixtures.

Concrete mix	Charge passed (C)					
	28 Days	% Variation with respect to RC 1	% Variation with respect to RC 2	56 Days	% Variation with respect to RC 1	% Variation with respect to RC 2
RC 1	704	-	-	683	-	-
RC 2	1053	+49.57	-	1050	+53.73	-
STRCA 25	700	−00.56	−33.52	680	−00.43	−35.23
STRCA 50	736	+04.54	−30.10	723	+05.85	−31.14
STRCA 75	787	+11.78	−25.26	796	+16.54	−24.19
STRCA 100	864	+22.72	−17.94	879	+28.69	−16.28

Note: + Sign represents increase in charge passed and – sign represents decrease in charge passed.
Abbreviation: STRCA, surface-treated recycled coarse aggregates.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

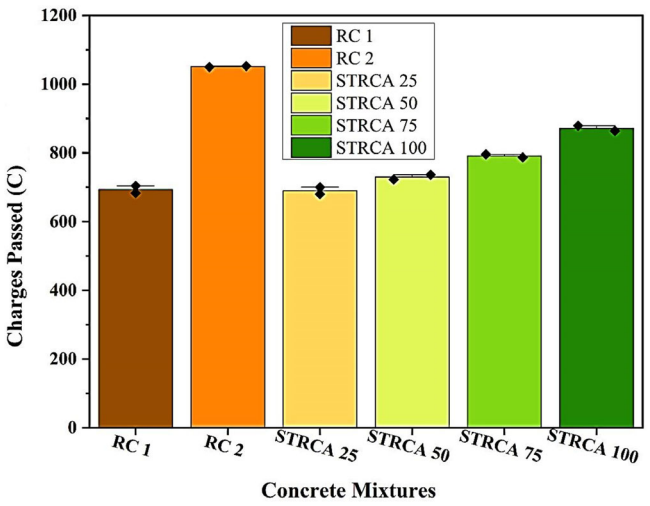


FIGURE 11 Variations in the amount of charges passed for different concrete mixtures. STRCA, surface-treated recycled coarse aggregates.

the control mix RC 2. The control mix RC 2 was produced using untreated RCA entirely substituted for NCA at a 100% replacement level.

In summary, the results indicate varied effects of STRCA on passed charge, with lower replacement levels (STRCA 25) showing marginal reductions and higher levels (STRCA 50, STRCA 75, and STRCA 100) leading to substantial increases. These findings suggest that surface modification crucially influences the chloride ion transport properties of concrete, with potential positive or negative impacts depending on the percentage of STRCA incorporated. Specifically, at a 100% replacement level, STRCA demonstrates a substantial 22%–28% reduction in chloride ion penetration compared to unmodified RCA. This aligns with prior research by Sasanipour and Aslani,[52] highlighting increased resistance with a higher water-cement ratio, and correlates with Zhan et al.'s[36] study on the impact of treated recycled aggregates on chloride permeability. The observed variations underscore the complex relationship between chloride ion transport and surface-modified recycled concrete aggregate, emphasizing the necessity for meticulous consideration of STRCA proportions in concrete mixtures. Despite diverse effects, the study indicates the potential for surface modification to be a valuable tool in tailoring concrete chloride resistance for specific applications. However, further research and optimization efforts are needed to precisely define optimal incorporation levels for desired chloride resistance properties. These findings underscore the importance of carefully considering the electrical properties of concrete when incorporating recycled materials, impacting the conductivity and overall electrical performance of concrete structures.

3.6 | Discussion

The abrasion treatment proved to be highly effective in addressing the surface of RCA, where it successfully dissolved the adhered mortar and introduced new pores, as evident in the SEM images (Figure 5) of ABTRCA. These alterations on the surface, credited to the cement slurry treatment, lead to the formation of a robust layer on the RCA surface. Notably, the utilization of cement slurry in treating RCA results in the filling of pores and cracks, consequently augmenting the strength of the treated RCA and yielding a smoother surface, as depicted in the SEM image (Figure 5) of CSTRCA. These modifications contribute significantly to the enhanced bonding with cement paste, thereby elevating the overall strength of the mixture. Despite these positive outcomes, the presence of capillary pores remains observable, and the hydrophilic nature of the hardened layer implies that water absorption does not decrease but may even experience a slight increase. When the STRCA is incorporated into RAC, there is a marked improvement in compressive strength and durability, attributed to the heightened strength and altered surface properties of the RCA. However, it is noteworthy that the water absorption of RAC is not noticeably affected, and the ITZ in RAC becomes more intricate. In cases where abrasion and cement slurry treatments are employed in conjunction, the comprehensive performance of STRCA sees further enhancement. This encompasses improvements in both strength and water absorption, ultimately leading to an augmented strength and durability of concrete incorporating STRCA.

4 | CONCLUSIONS

The study's conclusions propose that STRCA presents a viable and sustainable alternative for concrete production. Effective surface modification techniques mitigate the negative effects of contaminants on the surfaces of reinforced concrete, resulting in enhanced mechanical properties and increased durability. Upon conducting an in-depth analysis, a multitude of pivotal findings surfaces:

- Abrasion and cement slurry treatment efficiently enhance RCA by removing mortar and improving recycled aggregates, strengthening the interfacial connection with cement paste. Material characterization using XRD, SEM, and EDX reveals superior chemical properties in STRCA, showcasing positive enhancements in compound composition, intensity, and microstructure compared to conventional and unmodified counterparts.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

- Compared to the RC 1 mixture, STRCA 25 shows a significant 29.37% improvement in compressive strength at 25% replacement, but STRCA 100 experiences an 11.58% reduction at 100% replacement. STRCA can substitute up to 75% of coarse aggregates without a substantial loss in compressive strength. While improvements are seen up to a 50% replacement ratio, performance declines at 100% replacement.
- The drying shrinkage values of various concrete mixtures incorporating STRCA were notably lower than those of the reference concrete. At 100% replacement, STRCA demonstrates a reduction in shrinkage strain ranging from 11% to 20% compared to unmodified RCA.
- Surface modification techniques enhanced electrical resistivity, attributable to the dense microstructure of STRCA particles. At a 100% replacement level, STRCA demonstrates a 28%–36% increase in electrical resistance compared to unmodified RCA.
- STRCA concrete exhibits enhanced resistance to chloride ion penetration, classifying it as having low penetrability and a low risk of corrosion. At a 100% replacement level, STRCA demonstrates a significant 22%–28% reduction in chloride ion penetration compared to unmodified RCA.
- All STRCA integrated mixtures produced at various replacement levels (25%, 50%, 75%, and 100%) significantly outperformed the control mix RC 2, which was produced with untreated RCA substituted for NCA at a 100% replacement level.

This research affirms that the integrated surface modification of RCA, as demonstrated by the STRCA series, effectively enhances concrete durability. The applied surface modification techniques exhibit notable efficacy, resulting in concrete with improved durability characteristics. This strategy not only addresses challenges linked to recycled aggregates but also contributes to an overall advancement in concrete performance. The positive outcomes underscore the significance of the STRCA series as a promising avenue for improving concrete durability through innovative surface modification of recycled aggregates, making a valuable contribution to concrete technology.

5 | FUTURE SCOPE AND RECOMMENDATIONS

The positive outcomes of the STRCA series underscore its potential for concrete durability enhancement. Future research should focus on optimizing surface modification techniques, exploring alternative methods, and conducting in-depth microstructural analyses. Long-term durability studies, performance under varied environmental conditions, and scalability considerations are crucial for practical application. Integration with emerging technologies and a comprehensive life cycle assessment will further validate the sustainability of STRCA. Continued exploration of these aspects will contribute to advancing recycled aggregates in construction, promoting sustainable practices, and innovative solutions for the industry.

AUTHOR CONTRIBUTIONS

Harish Panghal: Conceptualization, methodology, formal analysis, investigation, resources, data curation, writing, original draft preparation, writing review and editing, and visualization. **Awadhesh Kumar**: Conceptualization, methodology, validation, resources, review and editing, supervision, project administration, and funding acquisition.

ACKNOWLEDGMENTS

The staff of Delhi Technological University's Department of Civil Engineering, located in Shahbad Daultpur Village, Rohini, New Delhi, is greatly appreciated by the authors for their invaluable support during the project.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable as all the data which were created or analysed in this study are already presented in form of figures and tables in the manuscript.

ORCID

Harish Panghal https://orcid.org/0000-0001-9930-4247

REFERENCES

1. Bansal S, Singh SK. Sustainable handling of construction and demolition (C & D) waste. Int J Sustain Energy Environ Res. 2015;4(2):22–48. https://doi.org/10.18488/journal.13/2015.4.2/13.2.22.48
2. Leopold M, Ntitanguranwa JP, Kabano J, Gasingwa N. Reuse of construction and demolished concrete waste by producing affordable high strength concrete block. Rwanda J Eng Sci Technol Environ. 2018;1(1):1–17. https://doi.org/10.4314/rjeste.v1i1.14s
3. Yu J, Liu D, Zhang Z. Durability and life prediction analysis of recycled aggregate concrete with ceramic waste powder under freeze-thaw conditions based on impact-echo method and Grey-Markov model. Front Mater. 2022;9(October):1–13. https://doi.org/10.3389/fmats.2022.1060294
4. Nanya CS, da Silva Ferreira FG, da Silva Capuzzo VM. Mechanical and durability properties of recycled aggregate concrete. Rev Mater. 2021;26(4):1–18. https://doi.org/10.1590/S1517-707620210004.1373
5. Diaz PM. Analysis of benefits, advantages and challenges of building information modelling in construction industry. J Adv Civ Eng. 2016;2(2):1–11. https://doi.org/10.18831/djcivil.org/2016021001
6. Aiyewalehinmi EO, Adeoye TE. Recycling of concrete waste material from construction demolition. Am J Eng Res. 2016;5(4):182–91.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7. Dardis JM. Recycling concrete for sustainable construction. ETD Archive. 2012;371. Available from: https://engagedscholarship.csuohio.edu/etdarchive
8. Hoffmann Sampaio C, Cazacliu B, Ambrós W, Kronbauer M, Tubino R, Dal Molin D, et al. Characterization of demolished concretes with three different strengths for recycling as coarse aggregate. Minerals. 2021;11(8):1–13. https://doi.org/10.3390/min11080803
9. Amasuomo E, Baird J. The concept of waste and waste management. J Manag Sustain. 2016;6(4):88. https://doi.org/10.5539/jms.v6n4p88
10. Kanungo J, Sood H, Student P. Effect of use of recycled coarse aggregate in concrete. Int J Eng Sci Comput. 2018;8(12):19479–83.
11. Jing Zhu B, Zheng W, Sneed LH, Xu C, Sun Y. Green demolition of reinforced concrete structures: review of research findings. Glob J Res Eng. 2019;19:1–19.
12. Safiuddin M, Alengaram UJ, Rahman MM, Salam MA, Jumaat MZ. Use of recycled concrete aggregate in concrete: a review. J Civ Eng Manag. 2013;19(6):796–810. https://doi.org/10.3846/13923730.2013.799093
13. Kalpavalli A, Naik SM. Use of demolished concrete wastes as coarse aggregates in high strength concrete production. Int J Eng Res Technol. 2015;4(7):1040–4.
14. Patel N, Patel P. Use of demolished concrete materials in concrete and comparative study of its mechanical properties: NDT comparison. Int J Innov Res Sci Eng Technol. 2016;3297(4):4712–21. https://doi.org/10.15680/IJIRSET.2016.0504015
15. Sonawane TR, Pimplikar SS. Use of recycled aggregate concrete. IOSR J Mech Civ Eng. 2013;2(1):52–9.
16. Phoeuk M, Kwon M. Accuracy prediction of compressive strength of concrete incorporating recycled aggregate using ensemble learning algorithms: multinational dataset. Adv Civ Eng. 2023;2023:1–23. https://doi.org/10.1155/2023/5076429
17. Zehra AFU, Sohoni VS. Development of modified concrete incorporating recycled aggregate and bagasse ash: a sustainable approach. Eur Chem Bull. 2023;12(4):194–200.
18. Pedro D, de Brito J, Evangelista L. Structural concrete with simultaneous incorporation of fine and coarse recycled concrete aggregates: mechanical, durability and long-term properties. Construct Build Mater. 2017;154:294–309. https://doi.org/10.1016/j.conbuildmat.2017.07.215
19. Tang Z, Li W, Ke G, Zhou JL, Tam VWY. Sulfate attack resistance of sustainable concrete incorporating various industrial solid wastes. J Clean Prod. 2019;218:810–22. https://doi.org/10.1016/j.jclepro.2019.01.337
20. Lal D, Dwivedi A, Chatterjee A. Microstructure study and strength behavior of cement mortar incorporating pond ash. J Emerg Technol Innov Res. 2018;5(6):64–71.
21. Mcginnis MJ, Gangone MV, Nogales A, Marisol L, Weldon GB, Reihl A. Experimental and numerical investigation of the bending, shear, and punching shear behavior of recycled aggregate concrete precast/prestressed hollow core slabs. Struct Concr. 2024;25:1–25. https://doi.org/10.1002/suco.202400008
22. Ulucan M, Gunduzalp E, Yildirim G, Alatas B, Alyamac KE. Optimizing sustainable concrete compressive strength prediction: a new particle swarm optimization-based metaheuristic approach to neural network modeling for circular economy and disaster resilience. Struct Concr. 2024;25:1–19. https://doi.org/10.1002/suco.202400070
23. Panghal H, Kumar A. Structural aspects of concrete incorporating recycled coarse aggregates from construction and demolished waste. Mater Constr. 2024;74(353):e337. https://doi.org/10.3989/mc.2024.360023
24. Panghal H, Kumar A. Examining the structural viability of recycled fine aggregates in sustainable concrete. J Mech Sci Technol. 2024;38(6):2931–42. https://doi.org/10.1007/s12206-024-04
25. de Juan MS, Gutiérrez PA. Study on the influence of attached mortar content on the properties of recycled concrete aggregate. Construct Build Mater. 2009;23(2):872–7. https://doi.org/10.1016/j.conbuildmat.2008.04.012
26. Wee TH, Suryavanshi AK, Tin SS. Influence of aggregate fraction in the mix on the reliability of the rapid chloride permeability test. Cem Concr Compos. 1999;21(1):59–72. https://doi.org/10.1016/S0958-9465(98)00039-0
27. Pandurangan K, Dayanithy A, Om Prakash S. Influence of treatment methods on the bond strength of recycled aggregate concrete. Construct Build Mater. 2016;120:212–21. https://doi.org/10.1016/j.conbuildmat.2016.05.093
28. Zhao Z, Remond S, Damidot D, Xu W. Influence of fine recycled concrete aggregates on the properties of mortars. Construct Build Mater. 2015;81:179–86. https://doi.org/10.1016/j.conbuildmat.2015.02.037
29. Li J, Xiao H, Zhou Y. Influence of coating recycled aggregate surface with pozzolanic powder on properties of recycled aggregate concrete. Construct Build Mater. 2009;23(3):1287–91. https://doi.org/10.1016/j.conbuildmat.2008.07.019
30. Ismail S, Ramli M. Mechanical strength and drying shrinkage properties of concrete containing treated coarse recycled concrete aggregates. Construct Build Mater. 2014;68:726–39. https://doi.org/10.1016/j.conbuildmat.2014.06.058
31. Chauhan BL, Singh GJ. Sustainable development of recycled concrete aggregate through optimized acid-mechanical treatment: a simplified approach. Construct Build Mater. 2023;399:132559. https://doi.org/10.1016/j.conbuildmat.2023.132559
32. Wang L, Wang J, Qian X, Chen P, Xu Y, Guo J. An environmentally friendly method to improve the quality of recycled concrete aggregates. Construct Build Mater. 2017;144:432–41. https://doi.org/10.1016/j.conbuildmat.2017.03.191
33. Alqarni AS, Abbas H, Al-Shwikh KM, Al-Salloum YA. Treatment of recycled concrete aggregate to enhance concrete performance. Construct Build Mater. 2021;307:124960. https://doi.org/10.1016/j.conbuildmat.2021.124960
34. Ahmad S, Maslehuddin M, Shameem M, Faysal RM, Adekunle SK. Effect of abrasion and chemical treatment of recycled aggregate on the workability, strength, and durability properties of concrete. Eur J Environ Civ Eng. 2022;26(8):3276–91. https://doi.org/10.1080/19648189.2020.1797886
35. Yunusa M, Zhang X, Cui P, Tian X. Durability of recycled concrete aggregates prepared with Mechanochemical and thermal treatment. Materials. 2022;15(16):5792. https://doi.org/10.3390/ma15165792
36. Zhan BJ, Xuan DX, Poon CS. Enhancement of recycled aggregate properties by accelerated CO_2 curing coupled with limewater soaking process. Cem Concr Compos. 2018;89:230–7. https://doi.org/10.1016/j.cemconcomp.2018.03.011

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

37. Quattrone M, Angulo SC, John VM. Energy and CO_2 from high performance recycled aggregate production. Resour Conserv Recycl. 2014;90:21–33. https://doi.org/10.1016/j.resconrec.2014.06.003
38. Shi C, Wu Z, Cao Z, Ling TC, Zheng J. Performance of mortar prepared with recycled concrete aggregate enhanced by CO_2 and pozzolan slurry. Cem Concr Compos. 2018;86:130–8. https://doi.org/10.1016/j.cemconcomp.2017.10.013
39. Feng C, Ye Z, Wang J, Wang L, Guo H, Zhang W, et al. Effect of cement slurry impregnation–carbonation strengthening on the properties of low-quality recycled concrete aggregate. Case Stud Constr Mater. 2024;20:e02914. https://doi.org/10.1016/j.cscm.2024.e02914
40. de Albuquerque e Silva CMM, Pereira MML, Capuzzo VMS, De Brito J. Concrete produced with recycled concrete aggregate exposed to treatment methods. Case Stud Constr Mater. 2022;18:e01938. https://doi.org/10.1016/j.cscm.2023.e01938
41. Jemimah Carmichael M, Prince Arulraj G. Rapid chloride permeability test on concrete with nano materials. Int J Eng Adv Technol. 2019;8(3):103–9.
42. Ashani HR, Mehta JV, Markna JH. Sem analysis of nano carbon black concrete. Int J Creat Res Thoughts. 2018;6(2):2320–882.
43. Shafiq N, Kumar R, Zahid M, Tufail RF. Effects of modified metakaolin using nano-silica on the mechanical properties and durability of concrete. Materials. 2019;12(14):1–22. https://doi.org/10.3390/ma12142291
44. Li L, Lu J, Shen P, Sun K, Pua LEL, Xiao J, et al. Roles of recycled fine aggregate and carbonated recycled fine aggregate in alkali-activated slag and glass powder mortar. Constr Build Mater. 2023;364:129876. https://doi.org/10.1016/j.conbuildmat.2022.129876
45. Qiu J, Tng DQS, Yang EH. Surface treatment of recycled concrete aggregates through microbial carbonate precipitation. Construct Build Mater. 2014;57:144–50. https://doi.org/10.1016/j.conbuildmat.2014.01.085
46. Amadi IG, Beushausen H, Alexander MG. Multi-technique approach to enhance the properties of fine recycled aggregate concrete. Front Mater. 2022;9:1–15. https://doi.org/10.3389/fmats.2022.893852
47. Yang S, Lee H. Drying shrinkage and rapid chloride penetration resistance of recycled aggregate concretes using cement paste dissociation agent. Materials. 2021;14(6):1478. https://doi.org/10.3390/ma14061478
48. Dhanya BS, Santhanam M. Performance evaluation of rapid chloride permeability test in concretes with supplementary cementitious materials. Mater Struct Constr. 2017;50(1):67. https://doi.org/10.1617/s11527-016-0940-3
49. Malakooti A. Investigation of concrete electrical resistivity as a performance based test. Master thesis. Logan: Utah State University. 2017.
50. Zhang J, Shi C, Li Y, Pan X, Poon C-S, Xie Z. Performance enhancement of recycled concrete aggregates through carbonation. J Mater Civ Eng. 2015;27(11):04015029-1–04015029-7. https://doi.org/10.1061/(asce)mt.1943-5533.0001296
51. Kessal O, Belagraa L, Noui A, Maafi N. Performance study of eco-concrete based on waste demolition as recycled aggregates. Mater Int. 2020;2(2):123–30. https://doi.org/10.33263/materials22.123130
52. Sasanipour H, Aslani F. Durability assessment of concrete containing surface pretreated coarse recycled concrete aggregates. Construct Build Mater. 2020;264:120203. https://doi.org/10.1016/j.conbuildmat.2020.120203
53. Han S, Zhao S, Lu D. Performance improvement of recycled concrete aggregates and their potential applications in infrastructure : A review. Buildings. 2023;13:1411. https://doi.org/10.3390/buildings13061411
54. Spaeth V, Djerbi Tegguer A. Improvement of recycled concrete aggregate properties by polymer treatments. Int J Sustain Built Environ. 2013;2(2):143–52. https://doi.org/10.1016/j.ijsbe.2014.03.003
55. Li Y, Zhang S, Wang R, Zhao Y, Men C. Effects of carbonation treatment on the crushing characteristics of recycled coarse aggregates. Construct Build Mater. 2019;201:408–20. https://doi.org/10.1016/j.conbuildmat.2018.12.158
56. Zhan BJ, Xuan DX, Zeng W, Poon CS. Carbonation treatment of recycled concrete aggregate: effect on transport properties and steel corrosion of recycled aggregate concrete. Cem Concr Compos. 2019;104:103360. https://doi.org/10.1016/j.cemconcomp.2019.103360
57. Gómez-Cano D, Arias-Jaramillo YP, Beraal-Conea R, Tobón JI. Effect of enhancement treatments applied to recycled concrete aggregates on concrete durability: a review. Mater Constr. 2023;73(349):e308. https://doi.org/10.3989/mc.2023.296522
58. Ismail S, Ramli M. Engineering properties of treated recycled concrete aggregate (RCA) for structural applications. Construct Build Mater. 2013;44:464–76. https://doi.org/10.1016/j.conbuildmat.2013.03.014
59. Kazemian F, Rooholamini H, Hassani A. Mechanical and fracture properties of concrete containing treated and untreated recycled concrete aggregates. Construct Build Mater. 2019;209:690–700. https://doi.org/10.1016/j.conbuildmat.2019.03.179
60. Tam VWY, Tam CM, Le KN. Removal of cement mortar remains from recycled aggregate using pre-soaking approaches. Resour Conserv Recycl. 2007;50(1):82–101. https://doi.org/10.1016/j.resconrec.2006.05.012
61. Dimitriou G, Savva P, Petrou MF. Enhancing mechanical and durability properties of recycled aggregate concrete. Construct Build Mater. 2018;158:228–35. https://doi.org/10.1016/j.conbuildmat.2017.09.137
62. Evangelista L, De Brito J. Durability of crushed fine recycled aggregate concrete assessed by permeability-related properties. Mag Concr Res. 2019;71(21):1142–50. https://doi.org/10.1680/jmacr.18.00093
63. Zheng Y, Zhang Y, Zhang P. Methods for improving the durability of recycled aggregate concrete: a review. J Mater Res Technol. 2021;15:6367–86. https://doi.org/10.1016/j.jmrt.2021.11.085
64. Panghal H, Kumar A. Effects of surface modified recycled coarse aggregates on concrete's mechanical characteristics. Mater Res Express. 2023;10(9):095506. https://doi.org/10.1088/2053-1591/acf915
65. Panghal H, Kumar A. Enhancing concrete performance : surface modification of recycled coarse aggregates for sustainable construction. Constr Build Mater. 2024;411:134432. https://doi.org/10.1016/j.conbuildmat.2023.134432
66. Shi C, Li Y, Zhang J, Li W, Chong L, Xie Z. Performance enhancement of recycled concrete aggregate – a review.

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

J Clean Prod. 2016;112:466–72. https://doi.org/10.1016/j.jclepro.2015.08.057
67. Wang R, Yu N, Li Y. Methods for improving the microstructure of recycled concrete aggregate: a review. Construct Build Mater. 2020;242:118164. https://doi.org/10.1016/j.conbuildmat.2020.118164
68. Bureau of Indian Standards. IS 269-2015, 'Ordinary Portland cement' Indian Standard, New Delhi. New Delhi: Bureau of Indian Standards; 2015.
69. Bureau of Indian Standards. IS 383-2016: Coarse and fine aggregate for concrete—specification. New Delhi: Bureau of Indian Standards; 2016.
70. Bureau of Indian Standards. IS 2386-193-Part 3: Methods of test for aggregates for concrete (no. reaffirmed 2021). New Delhi: Bureau of Indian Standards; 1974.
71. Bureau of Indian Standards. IS 2386-Part 4: Methods of test for aggregates for concrete (no. reaffirmed 2021). New Delhi: Bureau of Indian Standards; 2002.
72. Bureau of Indian Standards. IS 9103: Specification for concrete admixtures. New Delhi: Bureau of Indian Standards; 1999.
73. Bureau of Indian Standards. IS 10262-2009: Guidelines for concrete mix design proportioning. New Delhi: Bureau of Indian Standards; 2009.
74. Bureau of Indian Standards. IS 10262:2019: Concrete mix proportioning – guidelines (second revision). New Delhi: Bureau of Indian Standards; 2019.
75. Bureau of Indian Standards. IS 516-2020-Part 6: Determination of drying shrinkage and moisture movement of concrete samples. New Delhi: Bureau of Indian Standards; 2020. Available from: www.standardsbis.in
76. Polder R. RILEM TC 154 EMC: electrochemical techniques for measuring metallic corrosion. Mater Struct Constr. 2000;33:603–11.
77. ASTM C1202. Standard test method for electrical indication of concrete's ability to resist chloride ion penetration. Am Soc Test Mater. 2012;1–8. https://doi.org/10.1520/C1202-12.2
78. Wagih AM, El-Karmoty HZ, Ebid M, Okba SH. Recycled construction and demolition concrete waste as aggregate for structural concrete. HBRC J. 2013;9(3):193–200. https://doi.org/10.1016/j.hbrcj.2013.08.007
79. Kioumarsi M, Azarhomayun F, Haji M, Shekarchi M. Effect of shrinkage reducing admixture on drying shrinkage of concrete with different W/C ratios. Materials. 2020;13(24):1–16. https://doi.org/10.3390/ma13245721

AUTHOR BIOGRAPHIES



Harish Panghal, Department of Civil Engineering, Delhi Technological University, Delhi 1100042, India. Email: harish_phd2k18@dtu.ac.in.



Awadhesh Kumar, Department of Civil Engineering, Delhi Technological University, Delhi 1100042, India. Email: awadheshg@dtu.ac.in.

How to cite this article: Panghal H, Kumar A. Enhancing concrete durability and strength: An innovative approach integrating abrasion and cement slurry treatment for recycled coarse aggregates. Structural Concrete. 2024. https://doi.org/10.1002/suco.202400387

17517648, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/suco.202400387 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

arXiv:2407.17813v1 [cs.CV] 25 Jul 2024

Enhancing Model Performance: Another Approach to Vision-Language Instruction Tuning

Vedanshu[1*], M M Tripathi[1] and Bhavnesh Jaint[1]

[1]Electrical Engineering, Delhi Technological University 110042, Delhi, India.

*Corresponding author(s). E-mail(s): vedanshu_mt2k17@dtu.ac.in;
Contributing authors: mmtripathi@dce.ac.in; bhavneshjaint@dtu.ac.in;

Abstract

The integration of large language models (LLMs) with vision-language (VL) tasks has been a transformative development in the realm of artificial intelligence, highlighting the potential of LLMs as a versatile general-purpose chatbot. However, the current trend in this evolution focuses on the integration of vision and language to create models that can operate in more diverse and real-world contexts. We present a novel approach, termed Bottleneck Adapter, specifically crafted for enhancing the multimodal functionalities of these complex models, enabling joint optimization of the entire multimodal LLM framework through a process known as Multimodal Model Tuning (MMT). Our approach utilizes lightweight adapters to connect the image encoder and LLM without the need for large, complex neural networks. Unlike the conventional modular training schemes, our approach adopts an end-to-end optimization regime, which, when combined with the adapters, facilitates the joint optimization using a significantly smaller parameter set. Our method exhibits robust performance with 90.12% accuracy, outperforming both human-level performance (88.4%) and LaVIN-7B (89.41%).

Keywords: NLP, MLLM, Adapter

1 Introduction

In the landscape of modern computational linguistics, large language models (LLMs)[1–4] have been at the forefront, consistently expanding the boundaries of what machines understand about human language. This expansion has been largely driven by increases in the models' complexity, measured by the number of parameters, and

the breadth of pre-training data. One notable advancement in this domain is the concept of instruction tuning[1, 5–7], which has significantly enhanced the ability of LLMs to mimic human conversational patterns and effectively perform a variety of natural language processing (NLP) tasks, inching closer to the elusive goal of artificial general intelligence, as exemplified by models like GPT-3.5 [1]. The current trend in this evolution aims to imbue LLMs with multimodal understanding, particularly the integration of vision and language, to create models that can operate in more diverse and real-world contexts. This ambition has been partly realized in recent advancements like GPT-4 [8], which represents a leap forward by incorporating a substantial vision-language dataset directly into the training of a multimodal generative pre-trained transformer.

In the expansive domain of artificial intelligence, the integration of large language models (LLMs) with vision-language (VL) tasks has been a transformative development. However, this integration often comes with considerable computational and storage demands. Traditional parameter-efficient transfer learning (PETL) [9–13] methods, designed to reduce these demands by fine-tuning a minimal number of parameters, have not fully met the efficiency and performance needs, particularly for multimodal LLMs. Existing approaches also show a gap in performance compared to full model fine-tuning and exhibit significant redundancy when combining LLMs with vision models for tasks like image captioning[14] or text-to-image generation[15]. Moreover, modular training strategies, although beneficial, involve expensive VL pre-training and often necessitate extensive updating of LLM parameters, hindering quick and efficient model adaptation.

In light of these challenges, our model introduces an innovative approach that stands out from the current landscape. By not freezing the LLM and vision encoder and implementing these layers in a quantized format, our model significantly slashes the memory footprint. It's designed to support not just image-text but also text-only instructions, addressing a major limitation in models like BLIP2[16] and miniGPT4[17].

Our approach utilizes lightweight adapters to connect the image encoder and LLM, enabling joint optimization of image and language models without the need for large, complex neural networks. This framework allows for seamless shifting between single and multi-modal instructions without compromising natural language understanding capabilities. Unlike the conventional modular training schemes, our model adopts an end-to-end optimization regime, which, when combined with the lightweight adapters, facilitates the joint optimization of the entire multimodal LLM using a significantly smaller parameter set. This leads to an impressive reduction in storage overhead, outclassing existing solutions by a considerable margin.

Our method's ability to process both image-text and text-only instructions, along with its quantized form, ensures that it can be swiftly trained on a single GPU, marking a substantial leap in training efficiency. In summary, our key contributions are as follows:

- We present a novel method that bypasses the need for costly training while preserving the LLM's NLP capabilities, addressing a critical gap in current multimodal LLMs.

- Through rigorous experimentation, we demonstrate that not only does our model excel in efficiency, but also delivers competitive performance compared to existing multimodal LLMs, highlighting its potential as a versatile general-purpose chatbot.

2 Related Work

Parameter-Efficient Fine-Tuning (PEFT) techniques, in contrast to complete fine-tuning strategies, keep most of the parameters of pretrained models fixed while achieving similar performance on subsequent tasks. In the realm of PEFT, several methods have been investigated. These include prompt tuning[18–21], Low-Rank Adaptation (LoRA)[12], and adapters[9, 22, 23]. Prompt tuning entails adding trainable prompt tokens to pre-trained large models, either in input embeddings exclusively or throughout all intermediate layers. LoRA integrates trainable rank decomposition matrices into network weights, demonstrating promising fine-tuning capabilities, particularly in large generative models. Adapters incorporate lightweight adaptation modules into each layer of pre-trained transformers, providing flexibility across various domains. These PEFT approaches aim to find a middle ground between efficiency and effectiveness in fine-tuning pre-trained models for a variety of downstream tasks. LAVIN [24] examines the incorporation of lightweight adapters into transformer blocks. They propose the idea of Mixture-of-Modality Adaptation (MMA) to connect Language and Vision tasks (LLMs) with Vision-Language (VL) tasks, allowing for the concurrent improvement of image and language models. MMA utilizes a routing algorithm to facilitate the smooth transition of LLMs between single- and multi-modal instructions while maintaining their expertise in natural language understanding. However, this study argues that dynamic adapters such as MMA may not be essential for achieving similar levels of accuracy, which could streamline both time and space requirements.

3 Method

Our research introduces a distinctive approach, termed Bottleneck Adapter (BA), specifically crafted for enhancing large language models (LLMs) with vision-language capabilities. The Bottleneck Adapter seamlessly integrates into LLMs, bestowing them with multimodal functionalities. It adeptly navigates between single- and multi-modal instructions while allowing for a unified optimization of the entire multimodal LLM framework through a process known as Multi Model Tuning (MMT). This methodology is particularly noteworthy for its efficiency, significantly reducing both the computational resources and storage needed during training.

3.1 Bottleneck Adapter Structure

At the heart of our methodology lies the Bottleneck Adapter(BA). Traditional visual adapters typically include a non-linear function to bolster adaptation for NLP tasks. Our research, however, unveils that omitting this non-linearity does not detract from the Adapter's performance in visual tasks. As a result, we redefine the adapter function

$g(Z)$ as:

$$g(Z) = Z + \psi_u(\psi_d(Z)) \tag{1}$$

Here, ψ_u and ψ_d signify the dense projections that are standard in typical adapters. Observing that sparse transformations are fundamental in numerous vision modules, our design introduces a dense-to-sparse architecture for the Bottleneck Adapter. We articulate the transformation ψ_u as a group-wise operation:

$$\psi_u(Z) = [Z^{'}_{p0}W_{p0}, ..., Z^{'}_{pk}W_{pk}] + b \tag{2}$$

In this formulation, $Z_i \in \mathbb{R}^{nx\frac{c}{k}}$ represents the subset of features extracted from $Z \in \mathbb{R}^{mxc}$, with p denoting the total number of groups. $W_i \in \mathbb{R}^{\frac{c}{k}\times\frac{d}{k}}$ is the projection weight matrix, and $b \in \mathbb{R}^d$ is the bias term. This design, pivoting on sparse transformations, considerably trims the BA's complexity, rendering it more streamlined than the typical visual adapters.

3.2 Bottleneck Adapter Placement

Recognizing the profound disparities between visual and linguistic models, we meticulously consider the placement of adapters within the model architecture. Inspired by the LAVIN adapter placement strategy, we select to situate each adapter prior to the Transformer block. This strategic positioning is not limited to Multi-Head Attention (MHA), but is also extended to the Feed-Forward Network (FFN) within the Vision Transformer (ViT). Such a deliberate placement strategy ensures that our adapters are optimally positioned to promote effective and efficient integration and learning within the multimodal LLMs.

In essence, our Adapter-based methodology marks a significant step forward in the realm of vision language adaptation for LLMs, presenting a lightweight, resource-efficient, yet highly effective solution to augmenting the multimodal functionalities of these complex models.

3.3 Model

In our model architecture, we integrate the Bottleneck Adapter into a large language model (LLM), specifically LLaMA-2[25], and pair it with the diht-ViT[26] as the image encoder. For processing an input image $I \in \mathbb{R}^{h\times w\times 3}$, we extract the visual features from the [cls] tokens at every fourth layer of the ViT, represented as $V \in \mathbb{R}^{n\times d}$. In the image encoder architecture, adapters are strategically positioned before each multi-head attention module.

For text instructions, we utilize word embeddings, denoted as $T \in \mathbb{R}^{l\times c}$. To align the visual features with the LLM's dimensionality, we employ a visual adapter, which transforms V as follows:

$$V' = \rho(VW_v + b_v)W_t + b_t \tag{3}$$

Here, $W_v \in \mathbb{R}^{d\times d_v}$ and $W_t \in \mathbb{R}^{d_v\times c}$ are the transformation weight matrices, while $b_v \in \mathbb{R}^{d_v}$ and $b_t \in \mathbb{R}^c$ serve as bias terms. The activation function ρ is chosen as the SwiGLU[27] function for its efficiency. Importantly, the dimension d_v is selected to be considerably smaller than both d and c, optimizing resource usage.

The input to the LLM is defined as:

$$Z = \begin{cases} [u_m, V', T] & \text{for text-image input,} \\ [u_m, T] & \text{for text-only input.} \end{cases} \tag{4}$$

In this equation, $[\cdot]$ symbolizes concatenation. The LLM then predicts the next token in a step-by-step manner based on this multimodal input, which can be mathematically represented as:

$$p_t = \prod_{s=1}^{S+1} p(Q_s | Z, Q_{0:s-1}; \theta_m, \theta_r) \tag{5}$$

Here, $p_t \in \mathbb{R}^m$ indicates the probabilities of the predicted word, with m being the length of the word embeddings. The parameters θ_m and θ_r correspond to those of the LLM and adaptation modules, respectively.

This architecture offers a significant simplification and reduction in complexity compared to previous models such as LLaVA[28]. For instance, the visual neck of our model is six times smaller than that of LLaVA, yet the performance between the two models remains comparable, demonstrating the efficiency and effectiveness of our design.

4 Experiments

In Section 4.1, we detail our approach's multimodal performance on the ScienceQA benchmark. An ablation study conducted on the ScienceQA validation set is presented in Section 4.2.

4.1 Experiment Setup

Datasets and tasks.

Models. In our approach, we utilize the Vision Transformer Large/14 (ViT-L/14-336PX)[29], pre-trained using the DIHT framework[26], to serve as the image encoding mechanism. This process involves the extraction of visual features through six [cls] tokens, each derived from every fourth layer within the ViT-L/14-336PX structure. For the language model component, we incorporate the LLaMA-2-7B variant. The visual processing section, referred to as the visual neck, is configured with a dimension of 128. Meanwhile, the Adapter is specified to have a dimension of 16, with the temperature settings adjusted to 10 for 7B.

Baselines. We compare our method with the recent parameter-efficient fine-tuning methods and adapters where a new adapter structure with down-projection, up-projection and non-linear function are inserted into the transformer (both vision and NLP) and only the parameters of this new module are updated. Our baseline mmodels for comparison include the LaVIN-7B lite-clip[24], the LaVIN-13B - clip[24], the LLaMA adapter[30], LLaVA[28], Chameleon[31], the MM-CoTLarge[32], InstructBLIP (FlanT5XXL)[33] and BLIP-2 (FlanT5XXL)[16].

Implementation details. We process the image by directly resizing to 224×224 without any augmentation. The AdamW[34] optimization algorithm is employed to

fine-tune our model, which undergoes a training regimen spanning 20 epochs, guided by an absolute learning rate. The training process is characterized by a batch size of 1, a learning rate of 0.009, and a weight decay factor of 0.02.

During the text generation phase, a top-p sampling strategy is adopted, utilizing a temperature setting of 0.1 and a top-p threshold of 0.75.

4.2 Results and Analysis

Table 1 Accuracy on ScienceQA test set.

	Subject			Context Modality			Grade		Average
Method	NAT	SOC	LAN	TXT	IMG	NO	G1-6	G7-12	
Human	90.23	84.97	87.48	89.60	87.50	88.10	91.59	82.42	88.40
LaVIN-7B	89.25	94.94	85.24	88.51	87.46	88.08	90.16	88.07	89.41
LaVIN-13B	90.32	94.38	87.73	89.44	87.65	90.31	91.19	89.26	90.50
BLIP-2	-	-	-	-	-	-	-	-	89.5
InstructBLIP	-	-	-	-	-	-	-	-	90.7
MM-CoT-Base	87.52	77.17	85.82	87.88	82.90	86.83	84.65	85.37	84.91
MM-CoT-Large	95.91	82	90.82	95.26	88.8	92.89	92.44	90.31	91.68
LLaVA	90.36	95.95	88	89.49	88	90.66	90.93	90.9	90.92
Chameleon	89.83	74.13	89.82	88.27	77.64	92.13	88.03	83.72	86.54
LLaMA-Adapter	84.37	88.3	84.36	83.72	80.32	86.9	85.83	85.83	85.19
7B-lite-diht-BA(ours)	90.28	92.46	87.91	89.25	87.01	90.38	90.68	89.12	**90.12**

Our 7B-lite-diht-BA demonstrates competitive or superior performance, especially in the SOC and NO context categories. It signifies the model's balanced capability across different subjects and contexts, affirming the effectiveness of our proposed adaptations and optimizations. The human benchmark is 88. 40% average accuracy, highlighting the challenging nature of the ScienceQA dataset and the impressive capability of AI models such as LaVIN-13B, MM-CoTLarge, and our LaVIN-7B-lite-diht in approaching human-level performance on complex and multimodal scientific question answering.

Despite the higher performance of the MM-CoT-Large model, it is important to consider the advantages of our model's lighter and smaller architecture. The MM-CoT-Large model outperformed the 7B-lite-dith-BA (ours) model, likely due to a larger parameter space, sophisticated chain-of-thought reasoning capabilities, and potentially more effective multimodal integration. In social sciences, our method achieves 92.46% accuracy, surpassing LaVIN-7B (94.34%) but falling short of LaVIN-13B (94.38%) and LLAVa (95.95%). This indicates a strong performance, albeit with room for improvement in comparison to the top-performing methods. Our method's accuracy in language-related questions stands at 87.91%, which is competitive against LaVIN-7B (85.24%) but not as high as LaVIN-13B (87.73%) or LLAVa (88%). Our method gave an accuracy of 89.25% in text modality, which is comparably better than LaVIN-7B (88.51%) and closely aligned with the human benchmark (89.60%). With 87.01% accuracy in image-based questions, our method outshines LaVIN-7B (87.46%) but

is marginally outperformed by human accuracy (87.50%). The 7B-lite-diht-BA(ours) achieves an average accuracy of 90.12% across all categories. This is a remarkable result that closely mirrors human performance (88.40%) and surpasses LaVIN-7B (89.41%) by a narrow margin.

Table 2 This table presents the performance of each configuration across various accuracy metrics, demonstrating how changes in the adapter's dimension and group settings affect model performance in different contexts.

Adapter Dim	Groups	Overall Acc
8	2	87.88
16	2	90.12
4	2	86.49
32	2	84.15
16	4	89.74

Table 2 compares performance across various metrics for each configuration. The adapter dimension of 16 with 2 groups stands out as the most effective, achieving the highest overall accuracy and performing well across most individual metrics. Interestingly, increasing the adapter dimension to 32 with 2 groups resulted in lower performance, highlighting the potential for overfitting or inefficient parameter use at higher dimensions.

Table 3 This table presents the performance of various adapters and backbones, demonstrating how changes in the adapter and backbones affect model performance.

S No	Model	Backbone	Adapter	Overall Acc
1	7B-lite-clip	LLaMA	Router	88.35
2	7B-lite-clip	LLaMA-2	Router	88.78
3	7B-lite-alip	LLaMA-2	Router	87.97
4	7B-lite-diht	LLaMA-2	Router	88.68
5	7B-lite-diht	LLaMA-2	Concat Router	87.62
6	7B-lite-diht	LLaMA-2	Concat Router	86.88
7	7B-lite-diht-loha-1	LLaMA-2	loha	63.10
8	7B-lite-diht-loha-2	LLaMA-2	loha	55.10
9	7B-lite-diht-loha-3	LLaMA-2	loha	49.68
10	7B-lite-diht-lora-rep	LLaMA-2	lora-repAdapter hybrid	85.81
11	7B-lite-diht	LLaMA-2	bottleneck with weight scales	84.01

From table 3, if we look at the impact of the change in the backbone (different version of LLAMA), it can be seen that LLAMA-2 gives a slight edge in overall accuracy. With LLAMA[2] and Rep-Adapter[24, 35] we obtained an accuracy of 88.35%, while simply changing the backbone to LLAMA-2[25] we got an accuracy of 88.78 %. The vision backbone used here was CLIP[36]. With this result in mind, consecutive experiments were done using LLAMA-2 as a LLM backbone.

Motivated by CSPNet[37], we did concatenation of features in adapter. We look at the experiments 5 and 6 in which concatenation in adapter was used, it can be

found that concatenation didn't increase the accuracy. For concatenation equation 1 was modified to concatenate the features as: $[\psi_1(Z), \psi_2(Z)]$. The idea was to preserve the features with each blocks.

Experiment 7,8 and 9 were done using LoHa[38] as adapter. In the first one, we didn't use any down scaling convolution. We just used two LoHa conv_1d. While adding these two convolutions we used dynamic weight factors same as in LAVIN[24]. This is formulated as follows:

$$adapter(Z) = \hat{w}_1 \cdot f(Z, W_1) + \hat{w}_2 \cdot f(Z, W_2) \tag{6}$$

where, $\hat{w}_1$ and $\hat{w}_2$ are the routing weights and W_1 and W_2 are the Hadamard product of two low-rank inner matrices, $W := (X_1 Y_1^T) \odot (X_2 Y_2^T)$. In the second experiment, we simply replaced conv_1d in LAVIN with LoHa conv_1d. In the third expeitment, we added silu activation function without any weight factors which can be formulated as follows:

$$g(Z) = \phi_{L1}(Silu(\phi_{L2}(Z))) + Z \tag{7}$$

where, ϕ_{L1} and ϕ_{L2} are the LoHa convolution 1D operators.

Lower accuracy with LoHa can be linked with higher number of trainable parameter added in the adapter. Since, other adapters are using very less number of trainable parameters, they achieve higher accuracy with less data.

In the next experiment (11), we tried an hybrid of LORA and Rep-Adapter. In this LORA was used in LLM backbone (LLAMA-2) and Rep-Adapter was used in the vision backbone (CLIP). The reason of this experiment was to see the relation between dynamic Rep-Adapter with the vision. But the accuracy was not able to surpass the baseline.

In the final experiment we tried our mainline adapter with dynamic weight factors, but the results were not promising. Which allowed us to come to the conclusion that dynamic weight factors are not helpful for multi-modal fine tuning.

While selecting vision backbone we did experiment with CLIP[36], ALIP[39] and DIHT[26]. With CLIP we got an accuracy of 88.78% and with ALIP we got an accuracy of 87.98% and with DIHT we got an accuracy of 88.68%. While CLIP and ALIP used an image size of 224x224, DIHT uses an image of 336x336. With this in mind we used DIHT as our vision backbone even though CLIP had a slight edge over DIHT.

5 Conclusion

In this paper, we introduced a novel approach to enhance large language models (LLMs) with multimodal capabilities, addressing the critical challenge of computational and storage demands in integrating vision-language tasks. Our method leverages lightweight adapters and quantized encoding layers to significantly reduce the memory footprint, facilitating swift training on a single GPU without compromising the model's natural language processing capabilities. By eschewing conventional modular training strategies for an end-to-end optimization regime, we achieved notable efficiency and performance improvements over existing solutions.

6 Declarations

6.1 Conflict of interest

The authors declare that they have no conflict of interests.

6.2 Data availability

The study relies on openly accessible datasets, which can be found at https://scienceqa.github.io/. These datasets are freely available to researchers, promoting transparency, reproducibility, and facilitating further examination of the research outcomes.

References

[1] OpenAI: Chatgpt (2023)

[2] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971 (2023)

[3] Le Scao, T., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A.S., Yvon, F., Gallé, M., et al.: Bloom: A 176b-parameter open-access multilingual language model (2022)

[4] Peng, B., Alcaide, E., Anthony, Q., Albalak, A., Arcadinho, S., Cao, H., Cheng, X., Chung, M., Grella, M., GV, K.K., et al.: Rwkv: Reinventing rnns for the transformer era. arXiv preprint arXiv:2305.13048 (2023)

[5] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., *et al.*: Training language models to follow instructions with human feedback. Advances in neural information processing systems **35**, 27730–27744 (2022)

[6] Wei, J., Bosma, M., Zhao, V.Y., Guu, K., Yu, A.W., Lester, B., Du, N., Dai, A.M., Le, Q.V.: Finetuned language models are zero-shot learners. arXiv preprint arXiv:2109.01652 (2021)

[7] Iyer, S., Lin, X.V., Pasunuru, R., Mihaylov, T., Simig, D., Yu, P., Shuster, K., Wang, T., Liu, Q., Koura, P.S., et al.: Opt-iml: Scaling language model instruction meta learning through the lens of generalization. arXiv preprint arXiv:2212.12017 (2022)

[8] OpenAI: Gpt-4 technical report (2023)

[9] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International Conference on Machine Learning, pp. 2790–2799 (2019). PMLR

[10] Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. arXiv preprint arXiv:2101.00190 (2021)

[11] Lu, Y., Liu, J., Zhang, Y., Liu, Y., Tian, X.: Prompt distribution learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5206–5215 (2022)

[12] Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685 (2021)

[13] Liu, X., Ji, K., Fu, Y., Tam, W.L., Du, Z., Yang, Z., Tang, J.: P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks. arXiv preprint arXiv:2110.07602 (2021)

[14] Li, J., Li, D., Xiong, C., Hoi, S.: Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In: International Conference on Machine Learning, pp. 12888–12900 (2022). PMLR

[15] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)

[16] Li, J., Li, D., Savarese, S., Hoi, S.: Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In: International Conference on Machine Learning, pp. 19730–19742 (2023). PMLR

[17] Zhu, D., Chen, J., Shen, X., Li, X., Elhoseiny, M.: Minigpt-4: Enhancing vision-language understanding with advanced large language models. arXiv preprint arXiv:2304.10592 (2023)

[18] Jia, M., Tang, L., Chen, B.-C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.-N.: Visual prompt tuning. In: European Conference on Computer Vision, pp. 709–727 (2022). Springer

[19] Zhang, Y., Zhou, K., Liu, Z.: Neural prompt search. arXiv preprint arXiv:2206.04673 (2022)

[20] Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Learning to prompt for vision-language models. International Journal of Computer Vision **130**(9), 2337–2348 (2022)

[21] Zhou, K., Yang, J., Loy, C.C., Liu, Z.: Conditional prompt learning for vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16816–16825 (2022)

[22] Chen, S., Ge, C., Tong, Z., Wang, J., Song, Y., Wang, J., Luo, P.: Adaptformer:

Adapting vision transformers for scalable visual recognition. Advances in Neural Information Processing Systems **35**, 16664–16678 (2022)

[23] Jie, S., Deng, Z.-H.: Convolutional bypasses are better vision transformer adapters. arXiv preprint arXiv:2207.07039 (2022)

[24] Luo, G., Zhou, Y., Ren, T., Chen, S., Sun, X., Ji, R.: Cheap and quick: Efficient vision-language instruction tuning for large language models. Advances in Neural Information Processing Systems **36** (2024)

[25] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., Bashlykov, N., Batra, S., Bhargava, P., Bhosale, S., et al.: Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288 (2023)

[26] Radenovic, F., Dubey, A., Kadian, A., Mihaylov, T., Vandenhende, S., Patel, Y., Wen, Y., Ramanathan, V., Mahajan, D.: Filtering, distillation, and hard negatives for vision-language pre-training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6967–6977 (2023)

[27] Shazeer, N.: Glu variants improve transformer. arXiv preprint arXiv:2002.05202 (2020)

[28] Liu, H., Li, C., Wu, Q., Lee, Y.J.: Visual instruction tuning. Advances in neural information processing systems **36** (2024)

[29] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[30] Zhang, R., Han, J., Liu, C., Gao, P., Zhou, A., Hu, X., Yan, S., Lu, P., Li, H., Qiao, Y.: Llama-adapter: Efficient fine-tuning of language models with zero-init attention. arXiv preprint arXiv:2303.16199 (2023)

[31] Lu, P., Peng, B., Cheng, H., Galley, M., Chang, K.-W., Wu, Y.N., Zhu, S.-C., Gao, J.: Chameleon: Plug-and-play compositional reasoning with large language models. Advances in Neural Information Processing Systems **36** (2024)

[32] Zhang, Z., Zhang, A., Li, M., Zhao, H., Karypis, G., Smola, A.: Multimodal chain-of-thought reasoning in language models. arXiv preprint arXiv:2302.00923 (2023)

[33] Dai, W., Li, J., Li, D., Tiong, A.M.H., Zhao, J., Wang, W., Li, B., Fung, P.N., Hoi, S.: Instructblip: Towards general-purpose vision-language models with instruction tuning. Advances in Neural Information Processing Systems **36** (2024)

[34] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint

arXiv:1711.05101 (2017)

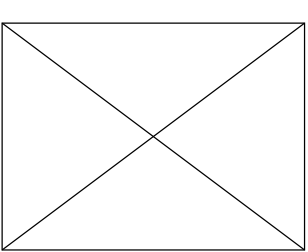
[35] Luo, G., Huang, M., Zhou, Y., Sun, X., Jiang, G., Wang, Z., Ji, R.: Towards efficient visual adaption via structural re-parameterization. arXiv preprint arXiv:2302.08106 (2023)

[36] Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., *et al.*: Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning, pp. 8748–8763 (2021). PMLR

[37] Wang, C.-Y., Liao, H.-Y.M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., Yeh, I.-H.: Cspnet: A new backbone that can enhance learning capability of cnn. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 390–391 (2020)

[38] Hyeon-Woo, N., Ye-Bin, M., Oh, T.-H.: Fedpara: Low-rank hadamard product for communication-efficient federated learning. arXiv preprint arXiv:2108.06098 (2021)

[39] Yang, K., Deng, J., An, X., Li, J., Feng, Z., Guo, J., Yang, J., Liu, T.: Alip: Adaptive language-image pre-training with synthetic caption. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2922–2931 (2023)



Energy Conversion and Management 316 (2024) 118823

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Evaluation of Statistical and Machine Learning Analysis on CRDI Engine using Ternary Fuel doped with Cerium Oxide Nanoadditives

Ashish Kumar Singh [a], Harveer Singh Pali [a,*], Nishant Kumar Singh [b], Sanjeev Kumar [c], Abhishek Sharma [d]

[a] Renewable Energy and Alternative Fuel Lab, Department of Mechanical Engineering, National Institute of Technology, Srinagar, Jammu and Kashmir, India
[b] Department of Mechanical Engineering, Harcourt Butler Technical University, Kanpur, Uttar Pradesh, India
[c] Department of Electrical Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, India
[d] Department of Mechanical Engineering, Loknayak Jai Prakash Institute of Technology, Chapra, Bihar, India

ARTICLE INFO

Keywords:
Biodiesel
Biobutanol
Cerium oxide nanoparticle
Machine learning
Optimization
Response surface methodology

ABSTRACT

At present, diesel engines are confronted with the formidable task of reconciling aggressive emission regulations with the preservation of intended engine performance. In order to tackle this concern, India can implement a comprehensive strategy that includes waste reduction, recycling programs, and the innovative use of waste-to-energy technologies. The potential synergistic effects of combining fuel mixtures with cerium oxide nanoparticles in CRDI diesel engines remain unexplored. This study employs response surface methodology and MLP regression to investigate these effects. This study explores the experimental design includes variation in blend ratios of karanja biodiesel (upto 20 %), compression ratios (18–22), and injection pressure (400–1200 bar), as well as CeO_2 nanoparticle concentrations (50–150 ppm) using 10 % biobutanol. A combination of response surface methodology (RSM) and artificial neural network (ANN) techniques is used to assess and optimise the influence of diverse process parameters on engine responses. The machine learning processor (MLP) regressor model is more capable of forecasting engine responses than RSM modelling, as it has lower MAE, RMSE, and higher R^2 values. According to the study, the best blend ratio, CR, Injection Pressure, and CeO_2 concentration for the best performance and lowest emissions are 9.97 % biodiesel blend, CR of 20, an injection pressure of 797.4 bar, and CeO_2 nanoparticle concentration of 100.1 ppm. The MAE value of the MLP regressor-based model ranged from 0.0011 to 0.0119. In terms of RMSE, MLP regressor-based model outputs are within reasonable boundaries. This study highlights the optimization of engine operating parameters and the application of nanotechnology in biodiesel blends. These findings can lead to the development of durable and eco-friendly fuel alternatives for internal combustion engines, promoting sustainable transportation.

1. Introduction

In the 21st century, there is a big issue regarding living organisms due to hazardous pollution. This harmful emission from vehicles and the automobile sector plays an important role in this problem of pollution. These problems have a significant impact on the very existence of human beings, and so must be addressed with the utmost care and caution [1]. In order to combat this problem, the Indian government has implemented various initiatives to reduce emissions from vehicles and to encourage the use of cleaner fuels. Additionally, the government has also implemented stricter emission standards for all vehicles, and stricter regulations for the use of conventional fuels like diesel, gasoline, etc. [2–5].

To achieve the aforementioned conditions, biodiesel has emerged as a viable alternative fuel, notably in India, which has a broad spectrum of flora and fauna [6]. However, due to the underpayment of workers

Abbreviations: ASTM, American Society for Testing and Materials; ANOVA, Analysis of Variance; BTE, Brake Thermal Efficiency; BSFC, Brake-Specific Fuel Consumption; CCD, Central Composite Design; CRDI, Common Rail Direct Injection; CO, Carbon Monoxide; CN, Cetane Number; CR, Compression Ratio; CV, Calorific Value; DoE, Design of Expert; EGR, Exhauts Gas Recirculation; FIP, Fuel Injection Pressure; HC, Hydrocarbon; IT, Injection Timing; ID, Ignition Delay; JB, Jatropha Biodiesel; KB, Karanja Biodiesel; KOH, Potassium Hydroxide; NOx, Nitrogen Oxide; RSM, Response Surface Methodology; VCR, Variable Compression Ratio.

* Corresponding author at: Renewable Energy and Alternative Fuel Lab, Department of Mechanical Engineering, National Institute of Technology, Srinagar, India.
E-mail address: hspali@nitsri.ac.in (H.S. Pali).

https://doi.org/10.1016/j.enconman.2024.118823
Received 11 May 2024; Received in revised form 18 July 2024; Accepted 18 July 2024
Available online 24 July 2024
0196-8904/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

involved in agriculture, a significant portion of the population remains impoverished. Researchers have conducted extensive investigations and concluded that India has the potential to meet its biodiesel production needs by carefully examining and exploring the plant species available in the country [7–10]. Non-edible oil crops such as Jatropha, Jojoba, Mahua, Polanga, Karanja, Simarouba, and others are gaining popularity as a technique of creating biodiesel without jeopardising food supply, which could lead to malnutrition, famine, and shortages. It is important to exercise caution in the production of biodiesel to ensure that it does not negatively impact food security [11,12]. Biodiesel is seen as more secure than diesel due to its higher flash and fire points. Because of its high oxygen content, its combustion produces less carbon monoxide and hydrocarbons (CO and HC), but it also increases emissions of NOx, one of the most harmful characteristics that can lead to acid rain and smog and thus ecological degradation. There is growing interest in undertaking further research to determine the best acceptable fuel option for contemporary diesel engines [13,14] in light of the existing energy catastrophes dependency on imported fossil fuels, and global warming. Ö. Cihan [15] discovered that biodiesel blends had lower levels of CO, particulate matter (PM), and UHC emissions. Higher biodiesel concentrations did, however, result in a modest rise in NOx emissions in comparison to diesel. According to Ho Young Kim et al. [16] the greater cetane number and oxygen percentage of biodiesel blends led to enhanced combustion efficiency. This resulted in decreased fuel consumption and improved thermal efficiency when weighed against pure diesel fuel. K Viswanathan et al. [17] investigated curry leaf biodiesel and discovered that biodiesel blends dramatically reduced CO, NOx, and total hydrocarbons (THC) emissions when compared to diesel. This demonstrates that biodiesel may serve as an environmentally preferable alternative to diesel fuel.Because of its low volatility and high viscosity, utilising pure biodiesel in diesel blends can lead to inadequate atomization and unfinished combustion, and it also creates carbon deposits [18].

Despite biodiesel's many benefits, there are some drawbacks to using it in internal combustion engines. Some biodiesel's much higher viscosity compared to diesel leads to issues including a smaller injection spray angle, deeper in-cylinder immersion of the fuel spritz, greater strain on fuel pump components, narrower poor vaporisation, huge droplet formation, and poor fuel atomization. In addition, biodiesel's mass-based gross energy density is roughly 12 % lower than diesel's, resulting in slower engine speed and less efficient operation. The increment of NOx emission is also a big issue with biodiesel [19–21]. The addition of nanoparticles to biodiesel and fitting with diesel, on the other hand, can relieve all of the difficulties connected with the direct use of biodiesel in internal combustion engines because it prevents significant engine alterations such as replacing the fuel line arrangement. Furthermore, the scientists suggest that a diesel–biodiesel combination with nanoparticles improves efficiency. In studies on emissions from this combination, researchers found low levels of CO and HC, as well as low levels of NOx and CO_2. As a result, it is prudent to investigate

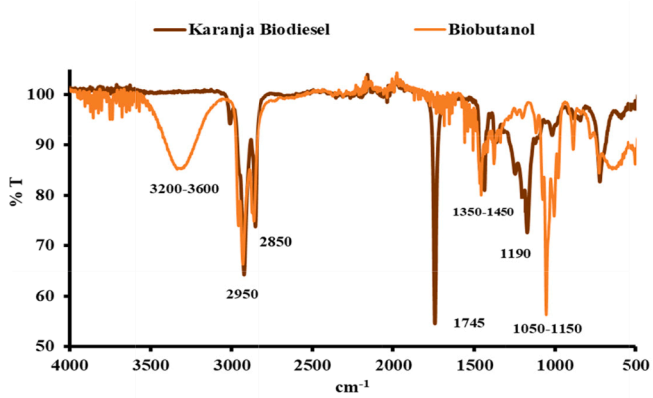


Fig. 2. FTIR of Karanja Biodiesel and Biobutanol.

nano-additives that can increase performance, reduce harmful emissions, and encourage greater fuel economy. There have been sevral studies [22–25] that looked at the potential benefits of using metal nanoparticle-based additives in biodiesel and its blends to boost engine performance. Metal oxides such as aluminium (Al_2O_3), titanium (TiO_2), copper (CuO), cerium oxide (CeO_2), carbon nanotubes (CNT), and more are known to help in the catalytic combustion of hydrocarbon fuels. As a consequence, introducing such ingredients to biodiesel fuels may enhance fuel properties including radiative mass transfer, ignition delay reduction, and self-ignition settings within the combustion zone [25–27]. Yusuf et al. [28] studied the effects of adding CeO_2 nanoparticles to biodiesel-diesel blends and discovered that all blends ofnanofuels enhance HRR, reduce CO and UHC emissions, and increase NOx emissions from 2.78 % to 19.01 %. Shengbo Ge et al. [29] revealed that TiO_2 nanoparticles at 50 ppm applied to blends (diesel + microalgae) enhance BTE by 5 %, reduce BSFC, raise HRR and cylinder pressure, and improve the combustion process. As a result, lower HC, CO, and CO_2, while increasing NOx.Cerium oxide (CeO_2) is utilised as an ingredient with diesel and biodiesel blends in this study to investigate its effect on the performance and emission aspects of CRDI engines. Due to its long-term viability, low cost, and simplicity of usage, the synthesis of CeO_2 has recently attracted a lot of attention. Therefore, CeO_2 was chosen as the preferred nano-additive for this study project. The oxygenated content of cerium oxide, which can enhance engine characteristics, is another factor in its selection as the chosen nano-additive for this investigation. Also, the addition of cerium oxide nanoparticles to biodiesel fuel enhanced its flash point and kinematic viscosity [30].

Numerous experts have undertaken comprehensive tests on diesel engines fueled with mixed diesel fuels, revealing a complex interplay of factors that influence engine performance. Singh et al. [31] delved into the impact of butanol blends on unmodified CI engines, noting a nuanced trade-off between engine power and brake thermal efficiency.

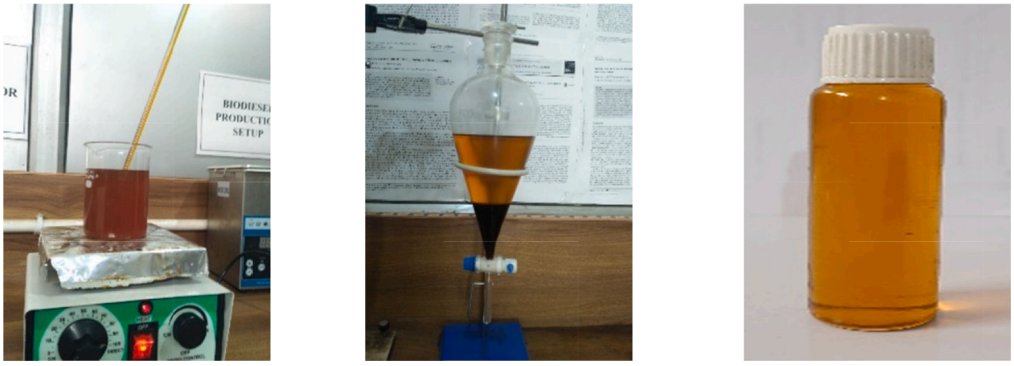
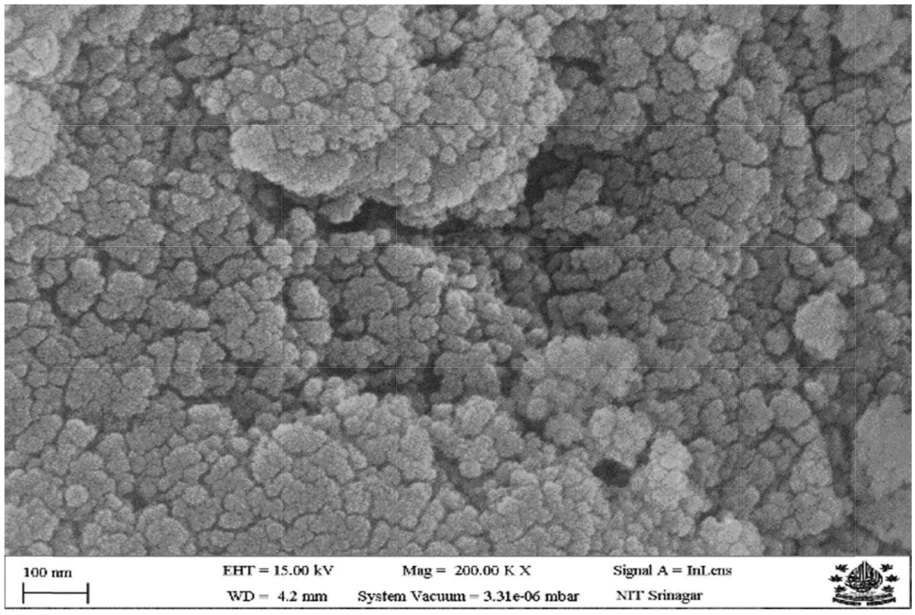
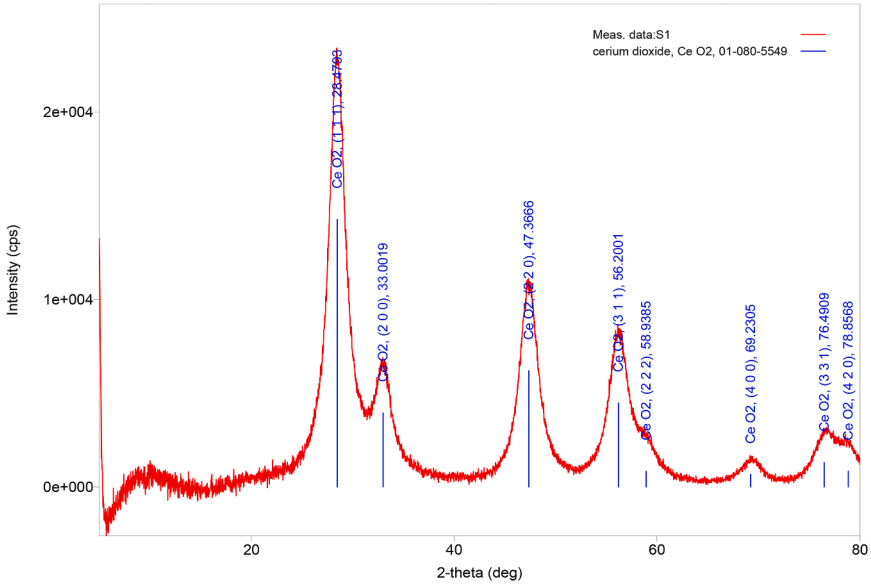


Fig. 1. The transesterification process to produce biodiesel from raw oil.



(a) SEM Image of CeO_2



(b) X-RD of CeO_2

Fig. 3. Characterisation of CeO_2 nanoparticles.

Meanwhile, Ge et. al.[32] shed light on the perplexing dynamics of particulate matter and emissions, highlighting the role of oxygen availability in mitigating pollutants while observing slight increases in certain emissions. Singh et. al. [33] contributed to the discourse by uncovering the surprising influence of oxygen on cold start combustion uncertainties, challenging conventional notions of fuel properties' importance. Yan et al. [34] explored the burstiness of optimal operational conditions in compression ignition engines with butanol diesel blends, emphasizing the significance of injection pressure thresholds. Parvesh et. al. [35] provided valuable insights into the burstiness of injection timing's effects, revealing a complex relationship between timing adjustments, emissions, and engine efficiency. Finally, Praveen et al. and Cai et al. [36,37] documented burstiness in brake thermal efficiency with varying oxygenate content, while also unveiling perplexing trends regarding ethanol's impact on emissions profiles. Together, these studies underscore the intricate nature of diesel fuel blends and their effects on engine performance, highlighting the importance of considering burstiness and perplexity in optimizing combustion processes.

The researchers had to undertake multiple trials to determine the optimal blend of elements for achieving the target results. As a result, researchers in numerous fields have begun to employ optimisation techniques to cut time consumption and forecast unknown combinations of parameters. The response surface method (RSM) is a mathematical tool for planning experiments and optimising variable effects [38]. The fundamental purpose of using RSM in experiment design is to maximise a response outcome (the contingent factor) that is influenced by controlled factors (the independent variables). The RSM optimisation

Table 1
Test Fuel Samples.

			
(D85KB05BB10)	**(D80KB10BB10)**	**(D75KB15BB10)**	**(D70KB20BB10)**
(85% D + 05% KB + 10% BB)	**(80% D + 10% KB + 10% BB)**	**(75% D + 15% KB + 10% BB)**	**(70% D + 20% KB + 10% BB)**

		
Diesel	**Biobutanol**	**Karanja Biodiesel**
(D)	**(BB)**	**(KB)**

Table 2
The physico-chemical characteristics of diesel, Karanja biodiesel, Biobutanol and Fuel Blends with ASTM standard.

Properties	ASTM Methods	Karanja Biodiesl (KB)	Biobutanol (BB)	Diesel (D)	D85KB05BB10	D80KB10BB10	D75KB15BB10	D70KB20B10
Density (kg/m^3)	D4052	872	810	840	838	840.5	842	843.5
Kinematics Viscosity (mm^2/s)	D445	4.712	3.1	2.92	3.02	3.1	3.2	3.31
Calorific value (kJ/kg)	D240	39,810	32,100	42,000	40,900	40,880	40,680	40,500
Flash Point (^{o}C)	D93	162	35	56	60	64	70	75
Cloud Point (^{o}C)	D2500	−3	−60	−13	−17	−16	−16	−15
Pour point (^{o}C)	D97	−4	−80	−15	−20	−20	−19	−19
Cetane number	D613	54	28	51	–	–	–	–
Sulphur content (ppm)	D4294	<10	<1	< 50	–	–	–	–
Acid Value (mg KOH/g)	D664	<0.5	<0.1	–	–	–	–	–

approach can also be utilised to improve the mixture ratio for binaries or ternary fuel in regard to engine performance, emissions, and nanoparticle concentration [39,40].

Unfortunately, several experimental research in the field of engineering necessitate large time and financial commitments. As a result, academics have focused their efforts on reducing the number of experimental trials [41]. The critical aspect in these endeavours is anticipating the experimental results. Tosunet et al. [42], predicted CO, NOx, and CO_2 emissions from a diesel and nanoparticle-blended CI engine. To forecast emissions, the scientists used ANFIS, regression analysis, and ANN algorithms in their study. According to their findings, the most successful model for estimating emissions in this investigation was ANFIS, subsequent to ANN and regression analysis. Machine learning models can complement complex physical models and optimize engine performance, emissions, and lifespan [43]. Meanwhile, Jafarmader et al. [44] used a wavelet neural network and a stochastic gradient technique to predict engine performance and emission attributes. During the algorithm's training phase, the authors took into account engine speed and nanoparticle percentages. They demonstrated unequivocally that the wavelet neural network was a fairly precise and efficient method for forecasting the nonlinear reactions of CI engines. A random forest machine learning tool, trained on experimental engine data, effectively predicts IMEP, reducing the need for extensive testing and aiding in engine design and optimization [45]. In another study, Saraee et al. [46] studied the impact of incorporating CeO_2 nanoparticles into a diesel engine on both performance and emission parameters. For projecting these traits, the authors also used the ANN algorithm. The results demonstrated that when forecasting CI engine reactions, ANN generated very pleasing results. Machine learning improves mixing models in high-Ka turbulent combustion by accurately specifying key parameters, significantly enhancing performance over empirical models with minimal sensitivity to input uncertainties [47].

As a result, researchers have demonstrated the capability of making very accurate predictions about experimental results. This innovation

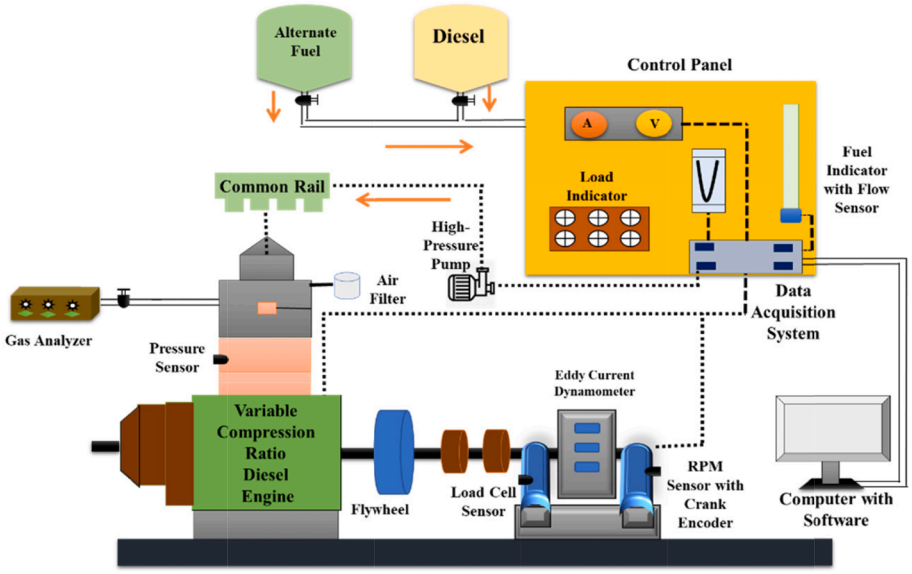


Fig. 4. The CRDI VCR diesel engine with eddy current dynamometer.

Table 3
Engine Setup and AVL Gas Analyzer Specification.

Type	Parameters	Specification
Diesel Engine	Make and Model	Kirloskar, CRDI VCR
	Power	3.5 kW at 1500 rpm
	Bore x Stroke	87.5 mm x 110 mm
	No. of Cylinder	1
	Cooling	Water cooled
Dynamometer	Working Principle	Eddy Current
	Speed	1200 – 2200 rpm
	Capacity	5 kW

Table 4
Specifiaction of AVL gas analyser and its range and accuracy.

Gas	Range	Accuracy	Measuring Principle
CO	0 – 10 %	± 0.02 %	Non-Dispersive Infrared (NDIR) Absorption
CO^2	0 – 20 %	± 0.03 %	Non-Dispersive Infrared (NDIR) Absorption
HC	0 – 10000 ppm	± 20 ppm	Flame Ionization Detection (FID)
NOx	0 – 5000 ppm	±10 ppm	Chemiluminescence Detection (CLD)

not only saves time and money [48,49] but also demonstrates the possibility for significant efficiency gains in the engineering area.

This study developed machine learning models to relate engine control variables to efficiency and emissions. The artificial neural network outperformed the random forest in accuracy for predicting combustion processes [50]. ANN outperformed other algorithms in accuracy, effectively removing noise and revealing that a higher compression ratio and near-stoichiometric equivalence ratio improve engine power without requiring compression ratio modifications [51].

The preceding research shows that fuel with nanoparticles is an excellent combination for enhancing engine performance and reducing emissions. The impacts of various process parameters, including applied load, engine speed (RPM), compression ratio, injection timing, injection pressure, fuel blends, additives, and nanoparticles, have been documented in the literature review [52,53]. Among these, fuel blends, injection pressure, doping of nanoparticles, and compression ratio have been identified as particularly significant for engine performance, especially in CRDI engines. Controlling of these parameters simultaneously makes polynomial response equations (parameters) almost difficult to address optimum response. So, new machine learning techniques like MLP regressor of ANN add novel dimensions to find the optimum solution with the help of RSM and, at the same time, CRDI engines with the provision of variable compression ratio and variable injection pressure provide unique response outcomes for transport engines. Specifically, the proper selection and optimization of parameters can substantially enhance engine performance without engine modifications.

Furthermore, it has been observed that other machine learning methods, such as the multilayer perceptron (MLP) regressor of ANN, are extensively used in a variety of technical domains. To the best of the Authors' knowledge, no previous work has used multilayer perceptron (MLP) regressor algorithms to forecast engine emission and performance parameters. To fill this void, the current work will apply this method to the emission and performance features of a diesel engine fueled by cerium oxide nanoparticles biodiesel-diesel blends.

1.1. Novelty

Previous research are concentrated on either biodiesel feedstocks or alcohols, but this study investigates the combination of biodiesels generated from Karanja and Biobutanol to improve fuel qualities and engine performance. The synergistic effects of combining fuel blends (karanja biodiesel, diesel and 10 % biobutanol) with cerium oxide nanoparticles are investigated in this work, which is largely unexplored in the context of CRDI diesel engines. Another distinctive feature of this study is the utilisation of the response surface approach and MLP regressor. RSM allows for the simultaneous optimisation of input variables while taking into account their interactions and nonlinear effects. RSM is utilize to determine the appropriate blending ratios of Karanja biodiesel, nanoparticle concentrations with variations of compression ratio and injection pressure. It resulted the best balance between engine performance (BSEC and BTE) and emissions characteristics (CO, NOx and HC). This study coupled with multilayer perceptron (MLP) regressor algorithms in ANN approch to forecast engine emissions and performance. The evaluation is based on two key metrics: MAE and RMSE. The findings should contribute to the optimisation of alternative fuels and nanotechnology in CRDI diesel engines, paving the way for more sustainable and efficient transportation systems.

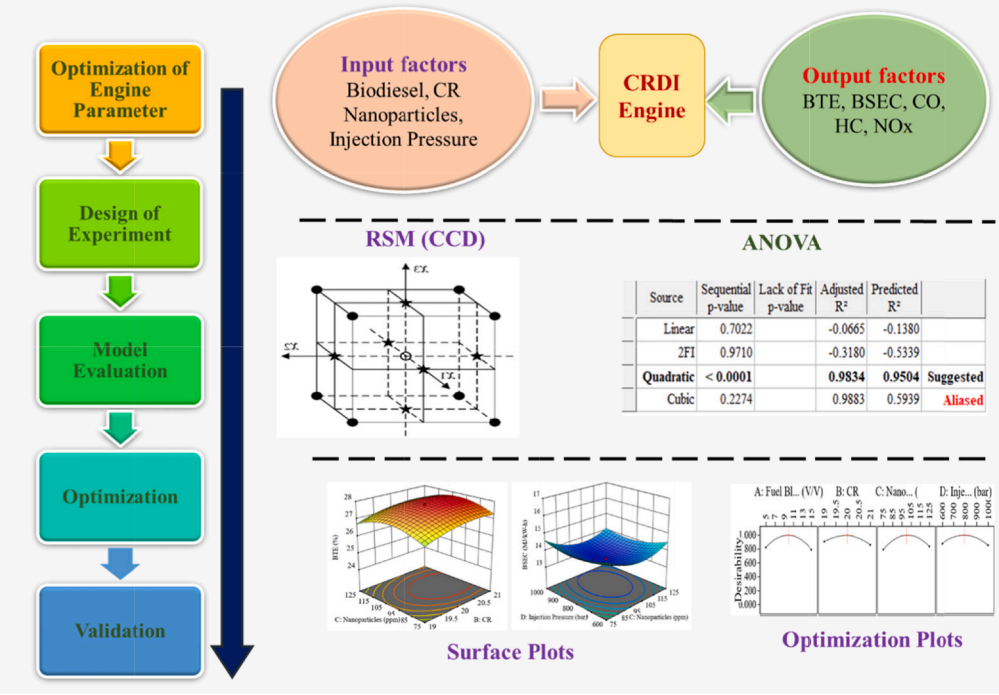


Fig. 5. Flow Chart of RSM.

Table 5
The different input attributes and their levels.

Input Parameters	Levels				
	−2	−1	0	1	2
Fuel Blends	0 (0 %KB+10 %B+D90)	5 (5 %KB+10 %B+D85)	10 (10 %KB+10 %B+D80)	15 (15 %KB+10 %B+D75)	20 (20 %KB+10 %B+D70)
CR	18	19	20	21	22
CeO_2 Nanoparticles (ppm)	50	75	100	125	150
Injection Pressure (%)	400	600	800	1000	1200

2. Materials and methods

2.1. Preparation of fuel

2.1.1. Biodiesel

Transesterification is a common method for manufacturing biodiesel from plant-based or fats from livestock. Step-by-step procedure for the transesterification process, which was developed specifically for manufacturing Karanja biodiesel: Karanja seeds are procured from botanical gardens of Delhi Technological University, and oil has been extracted through mechanical extraction processes. Oil has been refined through mechanical processes and is free from contaminants such as water, particles, and other pollutants. To begin, heat 100 ml of raw oil at 100 °C to remove water traces from the oil. Meanwhile, a sample of methoxide (Methanol and KOH in a 6:1 ratio) is created in a beaker. The raw oil is now being cooled to roughly 60 °C. The methanol-KOH combination was then progressively added to the heated Karanja oil in the reactor while stirring constantly to encourage excellent mixing and contact between the reactants. For about an hour, the reaction is kept at a temperature of 50–60 °C as shown in the Fig. 1. The catalyst promotes the conversion of triglycerides to methyl esters (biodiesel) by facilitating the interaction between Karanja oil and methanol. Allow the mixture to settle for a sufficient amount of time after the reaction to separate the glycerin phase (bottom layer) from the biodiesel (top layer). This settling stage could continue many hours or overnight. When the separation is complete, carefully drain the glycerin layer from the bottom of the reactor. The biodiesel was then properly cleansed with clean water to remove any remaining catalyst, soap, or impurities. Allow enough settling time after each wash and gently agitate the mixture while washing. The biodiesel was dried after washing to eliminate any leftover water. For purity of biodiesel Fourier Transform InfraRed (FTIR) of the karanja biodiesel is shown in Fig. 2, peak at 1745 cm^{-1} indicates the presence of ester carbonyl group, confirming the fatty acid methyl esters. The peak at 2850–2950 cm^{-1} corresponds to C–H methylene and methyl group in the fatty acid chains, whereas the peak at 1190 cm^{-1} is due to C-O stretching vibration confirming the ester linkage in biodiesel molecules.

2.1.2. Biobutanol

Fermentation of biomass source (sugar cane) is chosen to make Biobutanol of bio-origin. Biomass is introduced to clostridium acetobutylicum for the consumption of sugar and produces Biobutanol and byproducts like acetone and ethanol. After fermentation the mixture is separated using distillation, as Biobutanol has higher boiling point compared to ethanol and acetone. The crude Biobutanol is treated for solvent extraction for further refining to meet specific quality standards. For purity of biobutanol Fourier Transform InfraRed (FTIR) of the obtained biobutanol is shown in Fig. 2, a broad peak at 3200–3600 cm^{-1} indicates the O–H stretching vibrations, characteristics of alcohols. The peak at 2850–2950 cm^{-1} corresponds to C–H stretching in the alkyl group, whereas the peak in the range of 1350–1450 cm^{-1} indicates O–H bending vibrations, which further confirms presence of alcohol. Peaks around 1050–1150 cm^{-1} is due to C-O stretching vibration confirming the presence of alcohol functional group.

Table 6
The experimented array matrix.

Input Parameter					Responses				
S no.	**Blend (%)**	**CR**	**NP (ppm)**	**Injection Pressure (bar)**	**BTE (%)**	**BSEC (MJ/kW-h)**	**CO (%vol.)**	**HC (ppm)**	**NOx (ppm)**
1	10	20	100	800	27.8	13.5	0.16	21	532
2	10	20	100	800	27.8	13.5	0.16	21	532
3	5	21	125	1000	26.2	14.8	0.20	29	572
4	15	19	125	600	25.8	15.3	0.18	29	560
5	20	20	100	800	25.3	16.0	0.19	30	578
6	5	21	75	1000	26.4	15.1	0.20	30	570
7	10	20	50	800	25.8	16.5	0.19	31	566
8	15	19	75	600	25.4	15.6	0.18	29	567
9	15	21	125	600	26.3	15.1	0.18	28	572
10	10	20	150	800	25.6	15.5	0.19	28	562
11	0	20	100	800	25.5	15.8	0.19	29	559
12	10	20	100	800	27.8	13.5	0.16	21	532
13	15	19	125	1000	25.4	14.8	0.17	26	552
14	10	20	100	800	27.8	13.5	0.16	21	532
15	15	21	75	1000	26.2	14.8	0.21	30	580
16	10	20	100	800	27.8	13.5	0.16	21	532
17	5	19	125	1000	25.7	14.8	0.19	28	542
18	10	20	100	1200	26.2	14.8	0.19	27	558
19	15	21	125	1000	25.8	14.6	0.18	27	567
20	5	19	75	600	24.9	15.8	0.16	27	540
21	15	21	75	600	26.3	15.1	0.18	29	567
22	10	20	100	800	27.8	13.5	0.16	21	532
23	5	19	125	600	25.5	15.1	0.20	29	549
24	5	21	125	600	26.4	15.4	0.21	27	565
25	5	19	75	1000	26.0	15.5	0.19	29	551
26	10	20	100	400	25.8	15.2	0.18	26	552
27	15	19	75	1000	26.1	15.4	0.20	29	565
28	5	21	75	600	25.8	15.3	0.17	26	553
29	10	18	100	800	26.4	13.9	0.18	27	545
30	10	22	100	800	27.2	13.7	0.19	28	575

Table 7
The ANOVA analysis for BTE and BSEC.

BTE						BSEC					
Source	**Sum of Squares**	**df**	**Mean Square**	**F-value**	**p-value**	**Source**	**Sum of Squares**	**df**	**Mean Square**	**F-value**	**p-value**
Model	22.27	14	1.59	123.39	< 0.0001	**Model**	21.64	14	1.55	105	< 0.0001
A-Fuel Blends	0	1	0	0	1	A-Fuel Blends	0.0204	1	0.0204	1.39	0.2573
B-CR	1.6	1	1.6	124.27	< 0.0001	B-CR	0.2604	1	0.2604	17.69	0.0008
C-Nanoparticles	0.0067	1	0.0067	0.5172	0.4831	C-Nanoparticles	0.9204	1	0.9204	62.52	< 0.0001
D-Injection Pressure	0.2017	1	0.2017	15.65	0.0013	D-Injection Pressure	0.5704	1	0.5704	38.75	< 0.0001
AB	0.04	1	0.04	3.1	0.0985	AB	0.0506	1	0.0506	3.44	0.0835
AC	0.1225	1	0.1225	9.5	0.0076	AC	0.0156	1	0.0156	1.06	0.3192
AD	0.25	1	0.25	19.4	0.0005	AD	0.0006	1	0.0006	0.0425	0.8395
BC	0	1	0	0	1	BC	0.2256	1	0.2256	15.33	0.0014
BD	0.2025	1	0.2025	15.71	0.0012	BD	0.0056	1	0.0056	0.3821	0.5458
CD	0.64	1	0.64	49.66	< 0.0001	CD	0.0506	1	0.0506	3.44	0.0835
A^2	10.36	1	10.36	803.8	< 0.0001	A^2	9.77	1	9.77	663.74	< 0.0001
B^2	1.92	1	1.92	148.97	< 0.0001	B^2	0.1417	1	0.1417	9.62	0.0073
C^2	7.99	1	7.99	619.59	< 0.0001	C^2	10.61	1	10.61	720.5	< 0.0001
D^2	5.92	1	5.92	459.32	< 0.0001	D^2	3.79	1	3.79	257.65	< 0.0001
Residual	0.1933	15	0.0129			**Residual**	0.2208	15	0.0147		
Lack of Fit	0.1933	10	0.0193			Lack of Fit	0.2208	10	0.0221		
Pure Error	0	5	0			Pure Error	0	5	0		
Cor Total	22.46	29				**Cor Total**	21.86	29			

Table 8
The ANOVA analysis for CO, HC, and NOx emission.

CO						HC					
Source	**Sum of Squares**	**df**	**Mean Square**	**F-value**	**p-value**	**Source**	**Sum of Squares**	**df**	**Mean Square**	**F-value**	**p-value**
Model	0.0069	14	0.0005	55.82	< 0.0001	**Model**	295.13	14	21.08	189.73	< 0.0001
A-Fuel Blends	0.0001	1	0.0001	7.5	0.0152	A-Fuel Blends	0.6667	1	0.6667	6	0.0271
B-CR	0.0003	1	0.0003	30	< 0.0001	B-CR	0.1667	1	0.1667	1.5	0.2396
C-Nanoparticles	0	1	0	1.87	0.1911	C-Nanoparticles	6	1	6	54	< 0.0001
D-Injection Pressure	0.0004	1	0.0004	46.88	< 0.0001	D-Injection Pressure	1.5	1	1.5	13.5	0.0023
AB	0	1	0	2.81	0.1142	AB	0.25	1	0.25	2.25	0.1544
AC	0.0012	1	0.0012	137.81	< 0.0001	AC	4	1	4	36	< 0.0001
AD	0.00E+00	1	0.00E+00	0	1	AD	6.25	1	6.25	56.25	< 0.0001
BC	8.67E-19	1	8.67E-19	9.76E-14	1	BC	0.25	1	0.25	2.25	0.1544
BD	0	1	0	2.81	0.1142	BD	4	1	4	36	< 0.0001
CD	0.0012	1	0.0012	137.81	< 0.0001	CD	6.25	1	6.25	56.25	< 0.0001
A^2	0.0015	1	0.0015	173.57	< 0.0001	A^2	123.86	1	123.86	1114.71	< 0.0001
B^2	1.10E-03	1	1.10E-03	120.54	< 0.0001	B^2	72.43	1	72.43	651.86	< 0.0001
C^2	1.50E-03	1	1.50E-03	173.57	< 0.0001	C^2	123.86	1	123.86	1114.71	< 0.0001
D^2	0.0011	1	0.0011	120.54	< 0.0001	D^2	51.86	1	51.86	466.71	< 0.0001
Residual	0.0001	15	8.89E-06			**Residual**	1.67	15	0.1111		
Lack of Fit	0.0001	10	0			Lack of Fit	1.67	10	0.1667		
Pure Error	0	5	0			Pure Error	0	5	0		
Cor Total	0.0071	29				**Cor Total**	296.8	29			

NOx					
Source	**Sum of Squares**	**df**	**Mean Square**	**F-value**	**p-value**
Model	6910.88	14	493.63	197.89	< 0.0001
A-Fuel Blends	661.5	1	661.5	265.19	< 0.0001
B-CR	1350	1	1350	541.2	< 0.0001
C-Nanoparticles	20.17	1	20.17	8.08	0.0123
D-Injection Pressure	60.17	1	60.17	24.12	0.0002
AB	81	1	81	32.47	< 0.0001
AC	110.25	1	110.25	44.2	< 0.0001
AD	56.25	1	56.25	22.55	0.0003
BC	42.25	1	42.25	16.94	0.0009
BD	90.25	1	90.25	36.18	< 0.0001
CD	169	1	169	67.75	< 0.0001
A^2	2190.96	1	2190.96	878.34	< 0.0001
B^2	1272.96	1	1272.96	510.32	< 0.0001
C^2	1674.11	1	1674.11	671.13	< 0.0001
D^2	848.68	1	848.68	340.23	< 0.0001
Residual	37.42	15	2.49		
Lack of Fit	37.42	10	3.74		
Pure Error	0	5	0		
Cor Total	6948.3	29			

2.1.3. Cerium oxide nanoparticles

CeO_2 nanoparticles were purchased from an authorised vendor of Sigma Aldrich. Characterisation of the nanoparticles is shown in the Fig. 3. In Fig. 3 (a), the Scanning Electron Microscopy depicts that the average particle size of 25–30 nm and spherical structure of the particles. X-Ray Diffraction is illustrated in Fig. 3 (b) and detection of cerium oxide element in the nanoparticle confirming the purity of the nanoparticles.

Table 9
The regression model for output responses.

Parameters	S	R-sq	R-sq(adj.)	R-sq(pred.)
BTE (%)	0.1135	0.9914	0.9834	0.9504
BSEC (MJ/kW-h)	0.1213	0.9899	0.9805	0.9418
CO (vol.%)	0.0030	0.9812	0.9636	0.8915
HC (ppm)	0.3333	0.9944	0.9891	0.9677
NOx (ppm)	1.58	0.9946	0.9896	0.9690

2.2. Preparation of fuel samples

The test fuels were prepared by amalgamating karanja biodiesel, Biobutanol, and diesel alongside CeO_2 nanoparticles. The ratios of karanja biodiesel were varied up to 20 %, accompanied by 10 % Biobutanol, and CeO_2 nanoparticles ranging from 50 to 150 ppm within the fuel blends. These blends underwent ultrasonication at 40 Hz for a duration of half an hour. Following this procedure, it was noted that the samples remained steadfast for over 72 h without necessitating any surfactants.

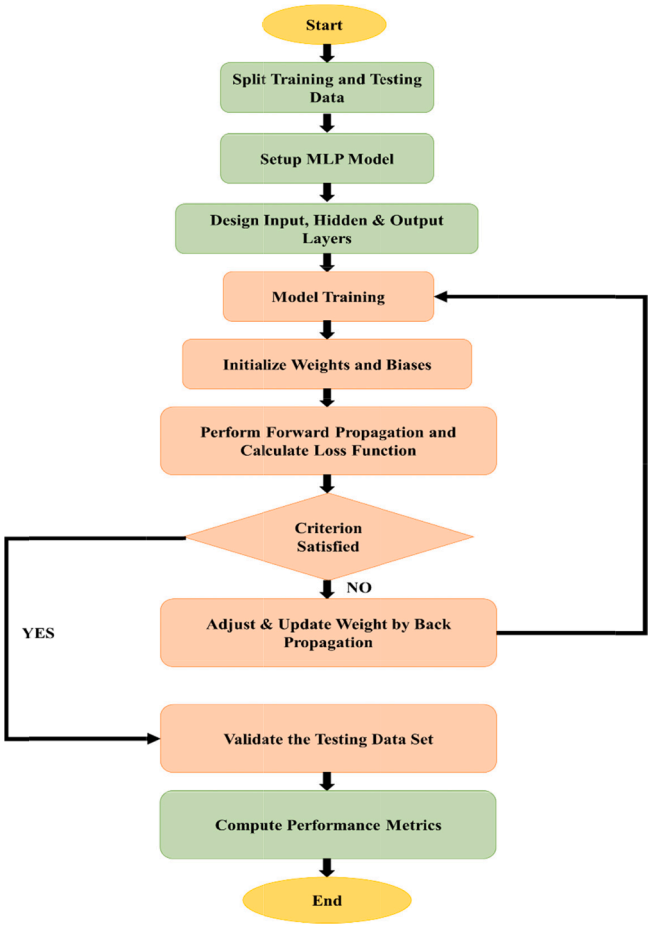


Fig. 6. Flow Chart of MLP algorithm.

Table 1 shows the test fuel samples of 100 ml in each bottle blended with diesel, Karanja biodiesel and biobutanol. The biobutanol ratio is kept constant at 10 %, and Karanja biodiesel is varied from 0 to 20 % v/v. where D refers to diesel, KB means Karanja Biodiesel and BB is Biobutanol. The term D85KB05BB10 means a blend of 85 % diesel, 5 % karanja biodiesel and 10 % biobutanol.

The physico-chemical properties of Karanja biodiesel and Biobutanol are measured and mentioned in Table 2 with ASTM standards.

Biobutanol, biodiesel-diesel blend, CeO_2 nanoparticles, and diesel biodiesel fuels had been employed as experimental fuels in this experiment. For the experiment, CeO_2 nanoparticles with a particle size range of 10 to 25 nm were utilised. The CeO_2 nanoparticles were mixed with diesel fuel at concentrations of 50, 75, 100, 125, and 150 ppm in fuel samples. The process used to disperse nanoparticles in the test fuel by using ultrasound reduces the possibility of nanoparticle aggregation for around 30 min.

2.3. Engine setup and procedure

All the testing has been done on partial full loading (i.e. 80 % load) with a corresponding torque of 24 Nm at an engine speed of 1500 rpm at around 4.8 bar of BMEP, on a single-cylinder 4-strokes, CRDI diesel engine of 661 cc with water cooling has been employed to conduct this experiment. Fig. 4 displays diagrammatic representations of the test rig and engine parameters. The engine's compression ratio can be adjusted by manipulating the position of the auxiliary piston using the screw rod and hand wheel assembly. The calibrated scale is provided on the screw rod assembly to enable exact positioning of the auxiliary piston to a different compression ratio (CR). The test engine was subjected to a load-through by utilising an eddy current dynamometer. The rotary encoders and piezoelectric pressure sensors were utilised to measure the combustion chamber pressure and crank angle. The weight loss of fuel in the fuel tank was determined by using a load cell, which allowed for the calculation of the rate of fuel flow. An airflow measurement was conducted using a manometer equipped with an aperture. The proximity sensor was utilised to determine the speed of the engine. Performance

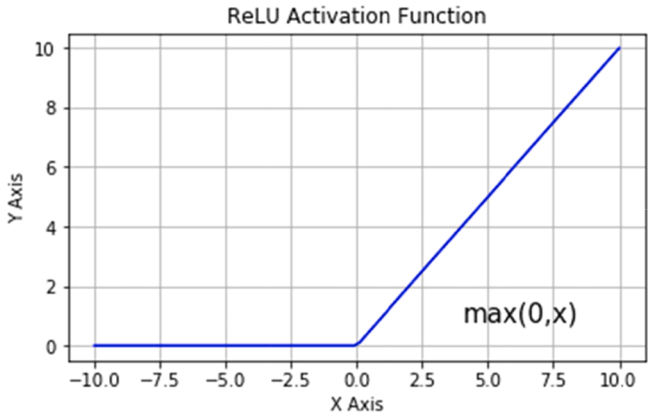


Fig. 8. Activation function of MLP regressor model.

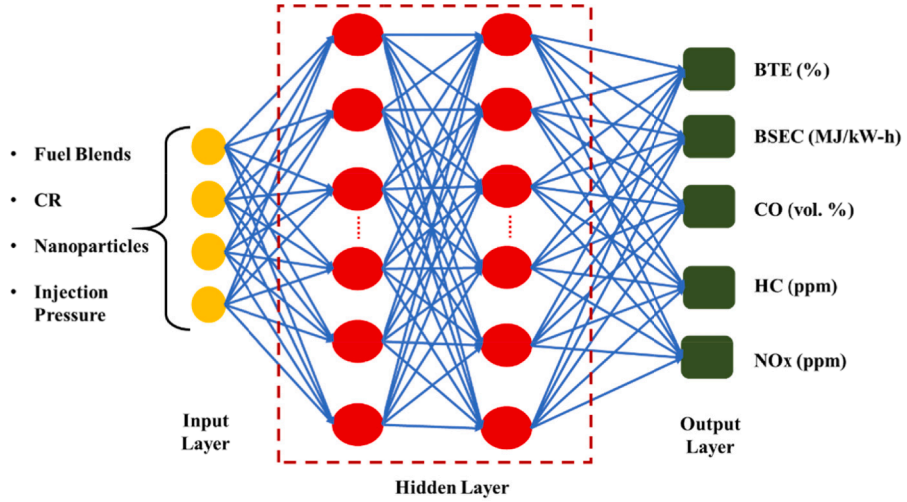


Fig. 7. Architecture of MLP Regressor Model.

```
[69] import pandas as pd
     import numpy as np
     from sklearn.model_selection import train_test_split
     from sklearn.neural_network import MLPRegressor
     from sklearn.preprocessing import StandardScaler
     from sklearn.preprocessing import MinMaxScaler
     from sklearn.metrics import mean_squared_error
     import matplotlib.pyplot as plt
```

```
     from google.colab import drive
     drive.mount('/content/drive')
```

```
[94] # Step 1: Load data from Excel file
     data = pd.read_excel('/content/drive/MyDrive/Karanja06March.xlsx')
     X = data[['input1', 'input2', 'input3', 'input4']]
     y = data[['output1']]
     #z = data[['r1', 'r2', 'r3', 'r4', 'r5']]
     z = data[['r1']]
```

```
[95] #  Normalize the data using Min-Max scaling
     scaler = MinMaxScaler()
```

Fig. 9. Steps of MLP regressor model.

parameters were evaluated subsequent to the collection and analysis of test parameters using the computerised engine analysis software "EngiSoft". The emissions of NOx, HC, and CO from the engine were monitored and analysed using an AVL gas analyzer. The specification of the engine setup and AVL Gas Analyser are given in Table 3 and Table 4.

2.4. Uncertainty analysis

Analysis of uncertainty is a crucial part of both engineering studies and scientific research. It entails evaluating and quantifying the degree of uncertainty around observations, computations, and model projections. Researchers can assess the correctness and dependability of their findings, pinpoint the causes of inaccuracy, and give a degree of confidence in their conclusions by doing an uncertainty analysis. Nevertheless, the uncertainty or error ratio is influenced by various factors, including the selection and calibration of the equipment, measurement methodologies, weather conditions, and experimental design. The Guide to the Expression of Uncertainity in Measurements (GUM) technique was used in this analysis to obtain the uncertainty, which is detailed below:

Total Uncertainty $= \pm \sqrt{\{}$(uncertainty of Speed sensor)^2 + (uncertainty of pressure sensor)^2 + (uncertainty of temperature sensor)^2 + (uncertainty of load cell)^2 + (uncertainty of Manometer)^2 + (uncertainty of fuel flow)^2 + (uncertainty of stop watch)^2 + (uncertainty of CO)^2 + (uncertainty of HC)^2 + (uncertainty of NOx)^2}.

$= \pm \sqrt{\{(0.1)^2 + (0.1)^2 + (0.15)^2 + (0.2)^2 + (1)^2 + (1)^2 + (0.2)^2 + (0.2)^2 + (0.2)^2 + (0.2)^2\}}$.

$= \pm 1.49$ %.

3. Optimization and Machine learning

3.1. Response surface methodology

Response Surface Methodology (RSM) is a statistical method for analysing and optimising systems or processes. Understanding the relationship between the input factors and the output factors of interest necessitates the use of mathematical models, specifically polynomial quadratic equations. RSM reduces the number of experiments needed and improves a response that is affected by several input variables. Fuel blends, CR, CeO_2 nanoparticle concentration, and injection pressure were all investigated in this study as input factors. Based on the prior research by Kumar et al. [52] and Pali et al.[54], the ranges for optimizing parameters have been determined, which were further refined through the conduction of pilot tests. BTE, BP, CO, HC, and NOx are a few examples of output reactions that these factors could affect. In this study a full factorial CCD is used which gives 30 runs in the design of experiments with 16 factorial points, 8 axial points and 6 center points. Flow chart of RSM is shown in Fig. 5. The input parameters and their levels are shown in Table 5.

The total number of experimental runs necessary to build the response surface when there are four variables with five levels each varies on the experimental design that is used. The Central Composite Design (CCD), which comprises factorial points, axial points, and centre points, is one often used design.The overall number of trials for a four-factor CCD with five levels is as follows:

Total number of experiments $=.2^p + 2p + c$

Where, 2^p is the factorial point, 2p is the axial point and c is the central point runs (i.e. 6), p is the number of input factor i.e. 4.

Total number of experiments $=2^4 + 2*4 + 6 = 30$ runs.

The 30 experiment runs were performed on the CRDI engine and their result were shown in Table 6.

3.2. RSM modelling and ANOVA analysis

The statistical metrics, such as R^2, Adjusted R^2, Predicted R^2, the F-value, and p-value, were calculated using ANOVA after conducting trials according to the design matrix. P-values below 0.05 are considered statistically significant when the F-value rises and the P-value drops, indicating greater significance. Table 7 presents the results of the engine performance, combustion, and emission reaction R^2, corrected R^2, and anticipated R^2. The second-order response surface models for brake thermal efficiency (BTE), brake specific energy consumption (BSEC), carbon monoxide (CO), hydrocarbons (HC), and nitrogen oxides (NOx) were constructed in Design Expert 13 using the experimental response data acquired from the central composite design (CCD) matrix. The BTE, BSFC, CO, HC, and NOx values are calculated using the equations provided in Tables 7 and 8. Table 9 presents the results of the engine performance and emission response, including the R^2, corrected R^2, and anticipated R^2 values.

3.3. Machine learning algorithms

Machine Learning (ML) is a common subset of Artificial Intelligence (AI) techniques. Because of its outstanding precision and efficacy, AI has received substantial attention and popularity due to its varied applications [46,48,49]. AI not only has an intrinsic attraction, but it also has an inherent potential to explore new realms and forecast uncharted scenarios. Without a doubt, the efficacy of ML algorithms is inextricably tied to user characteristics and the degree of success in their training. The performance of ML algorithms demonstrates their potential, which is moulded by user skill and the success of their training procedures. In the context of the current study, an in-depth examination of Multi-layer Perceptron Regressor (MLPR) algorithms was carried out. This algorithm was thoroughly explored, providing light on their unique traits, capabilities, and applications. A MLP regressor is preferred over other machine learning models due to its ability to capture complicated, nonlinear relationships in data, making it highly adaptable for a variety of regression applications. MLPs are capable of managing high-dimensional data and learning from detailed patterns, ensuring robustness against noisy data. Given a sufficient number of neurons in the hidden layers, they may approximate any continuous function, allowing for flexibility and power in a variety of regression applications. Furthermore, MLPs may be easily customized to different sorts of issues by modifying their architecture, which includes the number of layers and neurons, activation functions, and training procedures. MLPs also function easily with modern machine learning frameworks and can supplement other models in ensemble techniques, expanding their utility in a variety of applications.This analysis yielded a more detailed knowledge of the this algorithm, laying the groundwork for informed decision-making in their actual implementation. Fig. 6 give the flow diagram of MLP regressor.

Using MLPR algorithms to model engine responses, the study focuses on forecasting engine performance metrics (brake power and brake thermal efficiency) and emission characteristics (nitrogen oxides (NOx), carbon monoxide (CO), and hydrocarbons(HC). The algorithm uses four important input elements during the training phase: injection pressure (ranging from 400 to 1200 bar in 200 bar increments), biodiesel blend percentage (ranging from 0 to 20 %), CeO_2 nanoparticle concentration (ranging from 50 to 150 ppm), and compression ratio (18 to 22).A dataset of 30 data points is used to facilitate this investigation. The dataset is partitioned using a random shuffled sample strategy to maintain robustness. Notably, 70 % of the data points are set aside for algorithm training, with the remaining 30 % set aside for testing. The machine learning algorithm used in this study requires additional development. It uses complex quantitative methods to establish connections between the factors mentioned above and the engine's outputs (such as performance and emissions). Using this computational approach, we are able to comprehend the interplay between these numerous properties, which will lead to improvements in engine design and optimisation.

3.4. Multi-layer perceptron regressor (MLPR)

For many machine learning regression tasks, the multilayer perceptron (MLP) regressor is the ANN of preference. It is a feed forward neural network that can be trained to figure out intricate nonlinear connections between inputs and outputs, thanks to its many layers of artificial neurons (often called perceptrons).

To sklearn with a neural network. Scikit-learn's MLP Regressor represents an MLP regressor, a type of multi-layer perceptron. With its ability to do regression tasks, MLP is a sort of ANN. It may be trained to recognise non-linear associations between input and output.

Table 10
Metric equation, outline, and assessment.

Metrics	Equation	Description	Performance Criteria
MAE	$\mathbf{MAE} = \frac{\sum_{i=1}^{n}\left(\mathbf{Y_i^{predicted}} - \mathbf{Y_i^{experimental}}\right)^2}{\mathbf{n}}$	The MAE is a widely used indicator that provides useful insights into prediction models' prospective performance potential. Its widespread application emphasizes its importance in measuring predictive model efficacy over long time periods. MAE provides practitioners with a full grasp of a model's ability to maintain consistent accuracy across varied settings by quantifying the equilibrium between mistakes.	The lower value of MAE is a solid indicator that the used prediction model performs well. In fact, an MAE score of zero is the ideal circumstance, indicating a perfect match between forecasts and actual results.
RMSE	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i^{predicted} - Y_i^{experimental}\right)^2}{n}}$ Where, n = number of training data points; Y=values.	The RMSE is a valuable indicator of the prediction models' immediate efficacy. It is critical in determining how well these models correspond to real-world outcomes. RMSE provides insight into the accuracy and precision of forecasts by quantifying the difference between projected and observed values. This metric allows us to understand not only the existence of deviations, but also their amount, allowing us to conduct a thorough evaluation of the model's efficacy.	The RMSE is a measure that is constantly positive and should be as close to zero as possible. A lower RMSE suggests that a predictive model is more accurate in estimating values. In practice, an RMSE near to zero may indicate a significant alignment between anticipated and actual values, emphasizing the model's efficacy in minimizing prediction mistakes.
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{N}(X_i - Y_i)}{\sum_{i=1}^{N}(X_i - Y_k)}$ Where, $Y_k = \frac{\sum_{i=1}^{N} Y_i}{N}$ N=No training data X_i = Experimental data Y_i = Model anticipated data Y_k = Model average anticipated data	The coefficient of determination is a fundamental measure utilised to evaluate the quality of fit of a regression model.	Although the coefficient of determination offers valuable insights into the accuracy of a regression model, it is important to analyse it alongside other metrics and factors to make accurate inferences regarding the interaction between variables.

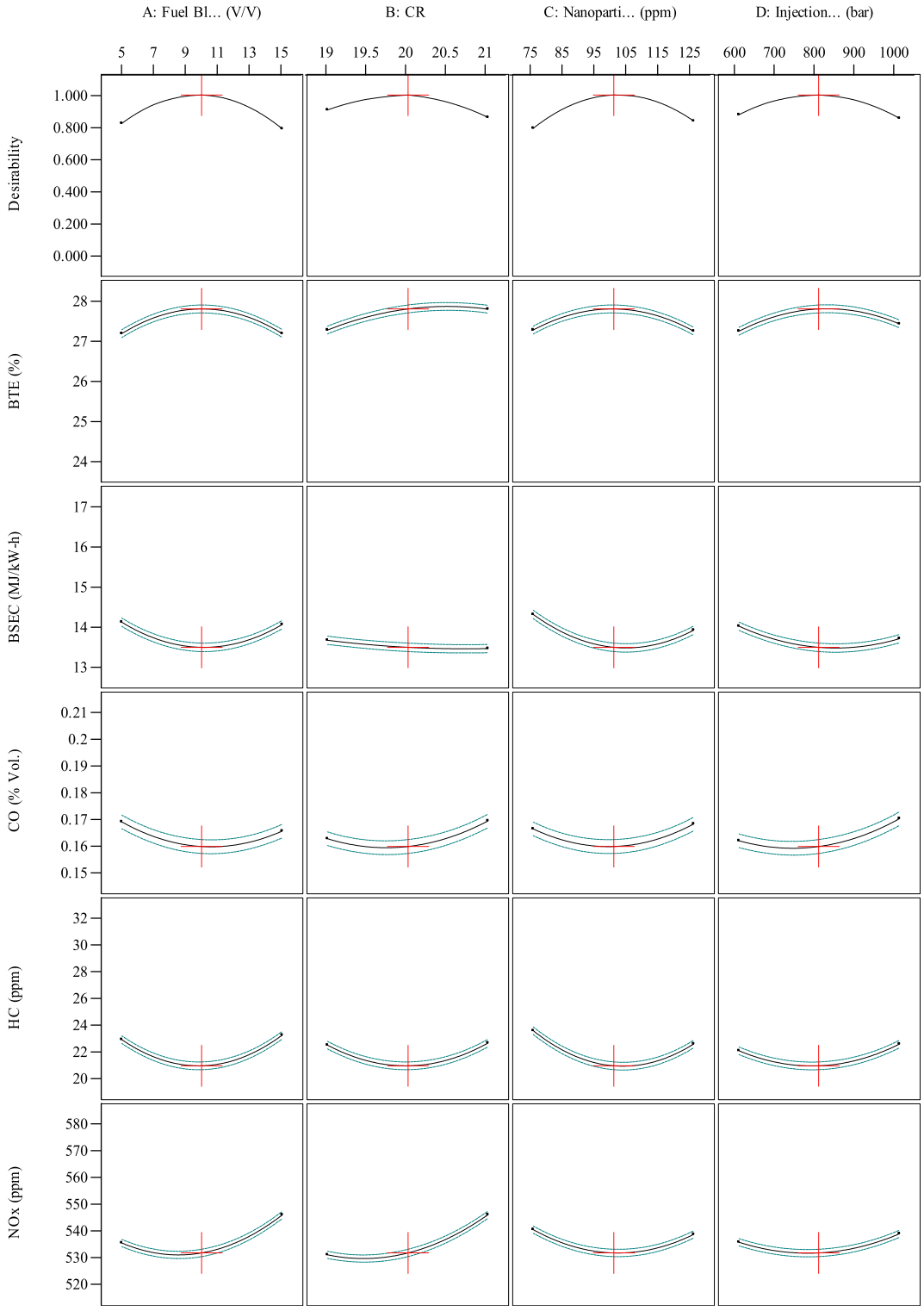


Fig. 10. Optimization plot.

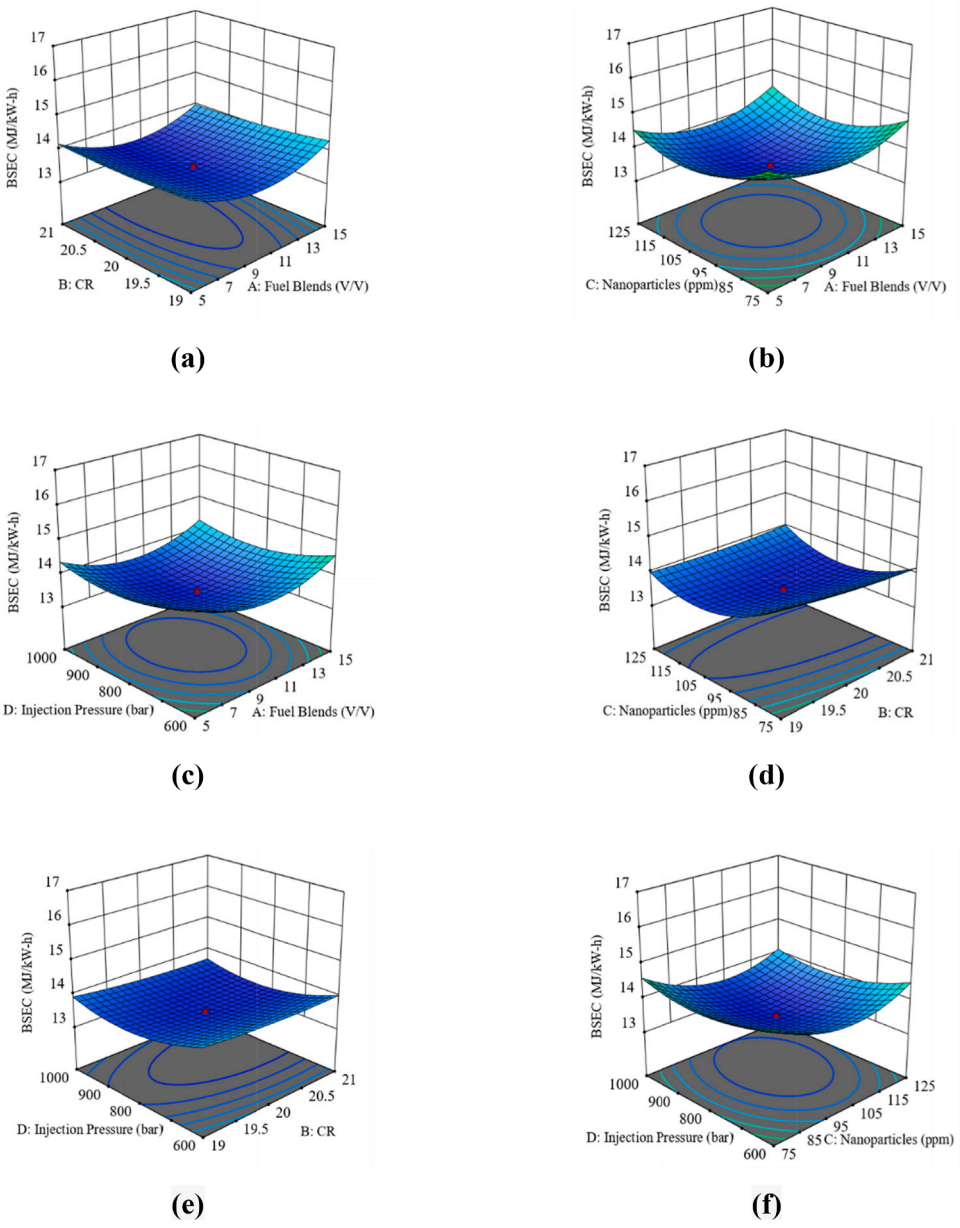


Fig. 11. Variation of BSEC against CR, fuel blend, injection pressure, and CeO_2 nanoparticles.

Throughout several industries, including finance, engineering, and healthcare, MLP regressors are used to address regression problems. Intricate relationships between inputs and outputs can be captured, and they can also be used to approximate complex nonlinear functions. To obtain optimal performance and avoid overfitting or underfitting, the number of hidden layers, the number of neurons in each layer, the activation functions, and the optimisation technique must be carefully chosen. An MLP regressor's architecture consists of three types of layers. The input layer gets four input characteristics (fuel mixes, CR, nanoparticles, and injection pressure), and each neuron represents a particular feature. The hidden layers, which sit between the input and output layers, are made up of many neurons that alter input data to learn complicated patterns. The output layer contains a single neuron that represents the projected output value (BTE, BSEC, CO, HC, NOx), and it generates the final predicted output.

The network architecture typically comprises of an input layer, 50 hidden layers, and an output layer when employing an MLP regressor for 4 inputs and 5 outputs. Each layer is made up of neurons, and the connections between them are represented by weights. Here's a full breakdown of the parts and stages needed in training an MLP regressor with four inputs and multiple outputs. The architecture of MLP regressor model is shown in Fig. 7. To bring nonlinearity into the network, activation functions are introduced to the output of each neuron in the hidden layers and the output layer. ReLU (Rectified Linear Unit), sigmoid, and hyperbolic tangent are examples of frequently employed activation functions. The activation function of MLP regressor model used in present work is shown in Fig. 8. The activation function in a MLP regressor model is an important component that affects the model's capacity to learn and predict. Activation functions add nonlinearity to the network, allowing it to learn from errors and create complicated mappings between inputs and outputs. Weight initialization is crucial to avoid the network being stuck in local optima during training. Before training, weights are generated randomly. During training, forward and backward propagation is performed on batches of training data and the weights are adjusted to minimize the loss function. The limited-memory Broyden-Fletcher-Goldfarb-Shanno (lbfgs) optimization method can be

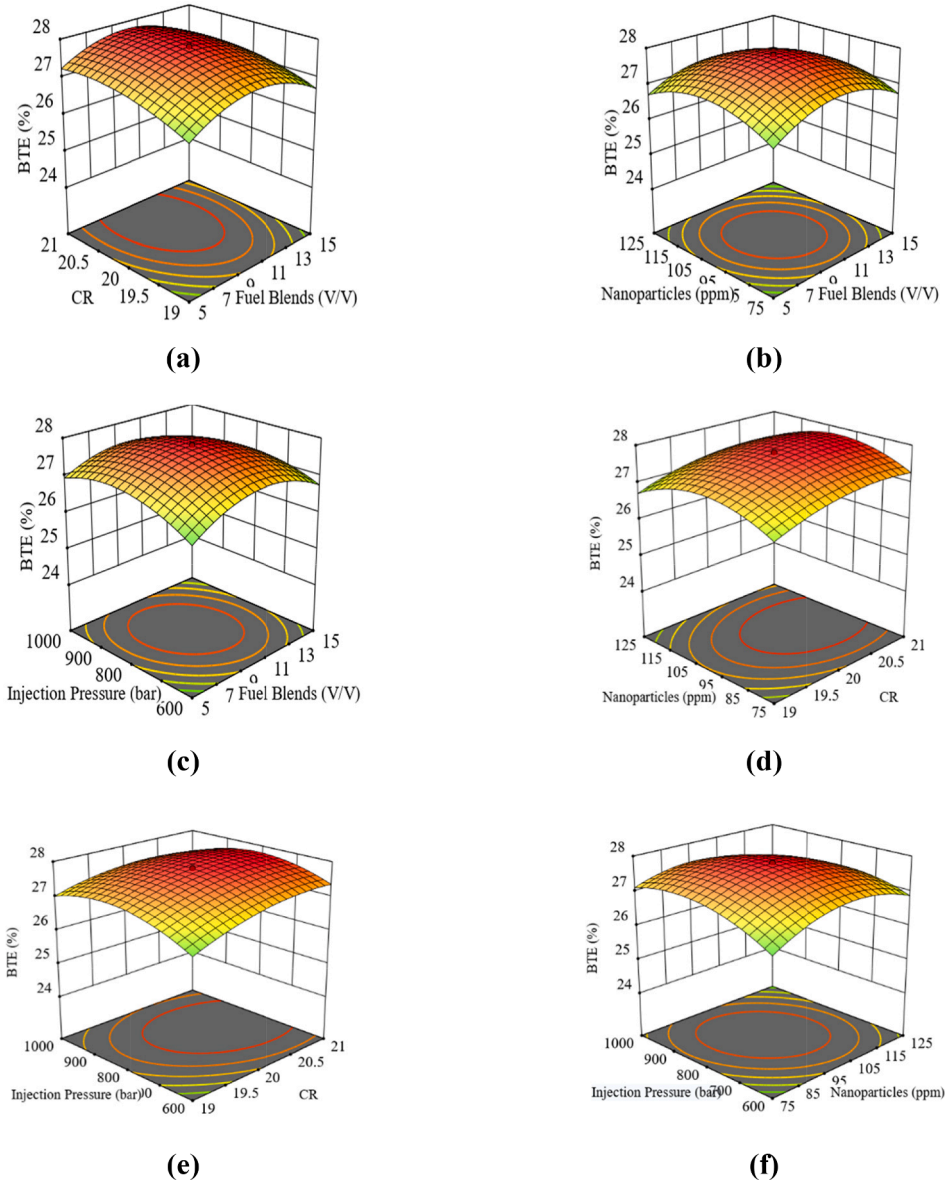


Fig. 12. Variation of BTE against CR, fuel blend, injection pressure, and CeO_2 nanoparticles.

used to adjust the weights during training. This optimizer belongs to the quasi-Newton method family and adjusts the learning rate and momentum to optimize the network's convergence.

3.5. Validation and Testing

Following training, the MLP regressor's performance is assessed on different validation and testing datasets. The regression metrics mean squared error (MSE), mean absolute error (MAE), and R-squared (R^2) are produced to assess the model's accuracy and generalisation ability. The Fig. 9 depicts a series of Python code executed in a Jupyter Notebook environment for the purpose of data processing and machine learning tasks. The starting code block includes the necessary libraries, such as pandas and numpy for data manipulation, sklearn for machine learning models and preprocessing, and matplotlib.pyplot for data visualisation. Subsequently, the code establishes a connection to Google Drive in order to get external files, emphasising the incorporation of cloud storage for data access. The dataset is imported from an Excel file, with certain columns chosen as inputs and outputs, indicating a setup suitable for regression analysis. The last code sample demonstrates data normalisation using the Min Max Scaler from the sklearn library. This process scales the characteristics of the data to a designated range in order to prepare it for modelling. This configuration exemplifies a standard machine learning procedure designed to train models using preprocessed data. The offered code presents a transparent and replicable approach for researchers to preprocess data, specifically by using cloud storage for data retrieval and implementing normalisation algorithms.

A variety of statistical criteria have been devised in the existing body of literature to analyse the performance of prediction models accurately. We used two widely recognized measures that are commonly used in this domain in the context of our study: MAE and RMSE. We hope to completely evaluate our models' prediction capability by incorporating these well-regarded indicators. RMSE provides a full grasp of the dispersion between anticipated values and actual observations, whereas MAE provides insights into the overall bias demonstrated by the model's predictions. Table 10 exhaustively documents the mathematical formulations of these indicators, as well as extensive descriptions and criteria for evaluating their efficacy. This detailed presentation ensures that everyone understands how these indicators contribute to the

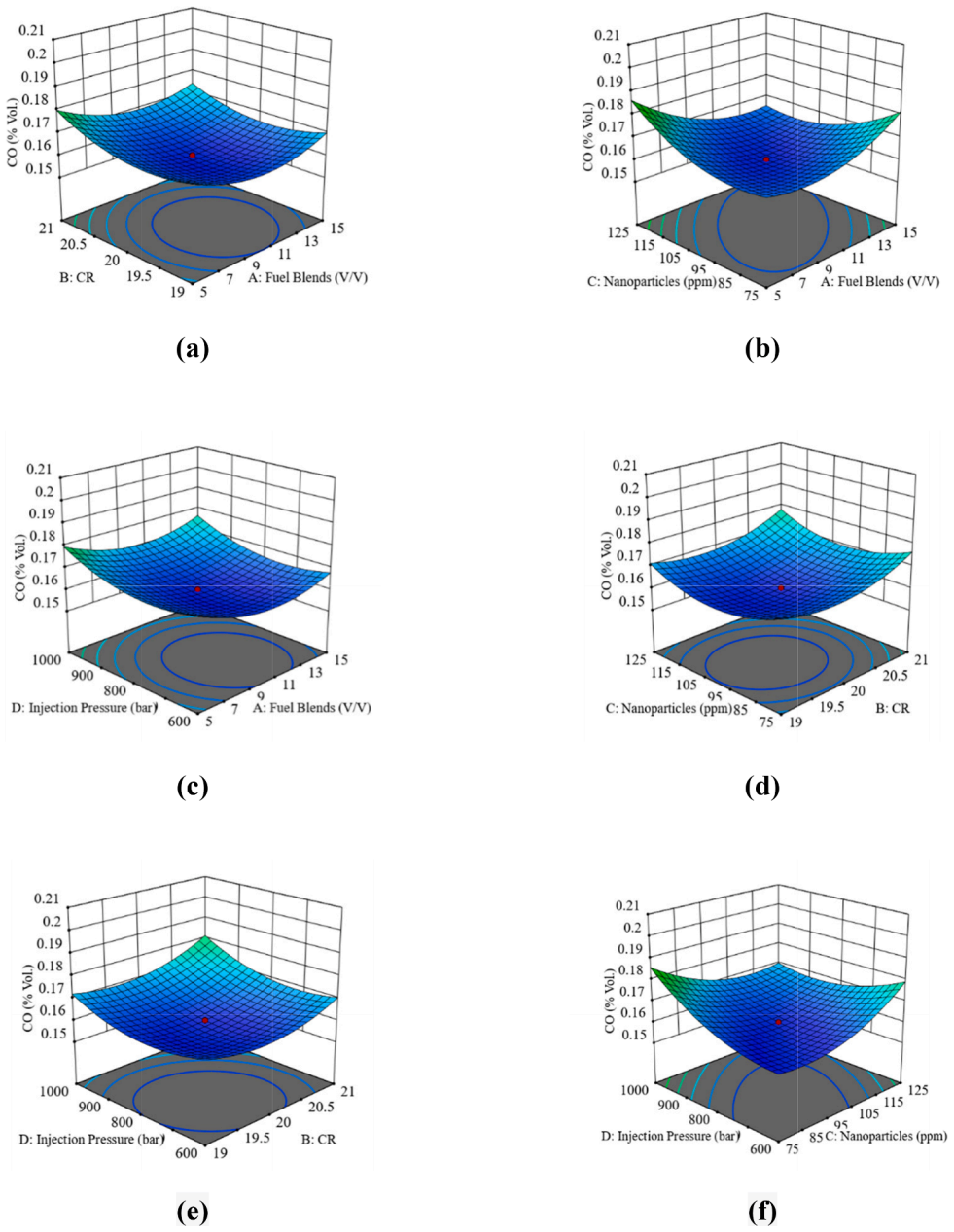


Fig. 13. Variation of CO emission against CR, fuel blend, injection pressure, and CeO_2 nanoparticles.

evaluation of our prediction models and helps us make relevant conclusions from our research.

4. Result and discussion

4.1. Results obtained from RSM

This section focuses on the effect of diesel engine input factors on five engine output responses. During surface plotting, the four-engine input variables (CR, Fuel Blend, CeO_2 concentration, and Injection Pressure) cannot be changed concurrently. Fig. 10 demonstrates the optimizer, and the optimised result for process parameters were obtained as fuel blend of 9.91 %, CR of 20.018, CeO_2 concentration of 10.696 ppm and injection pressure of 793.109 bar. At the optimal engine conditions are BTE of 27.8 %, BSEC of 13.5 MJ/kW-h, CO as 0.16 vol%, HC of 20.981 ppm, and NOx of 531.986 ppm. The obtained desirability of the sytem was found as 0.95.

4.1.1. Effect of control factors on brake-specific energy consumption

BSEC is an important metric for evaluating the efficiency of internal combustion engines. Lower BSEC values indicate greater energy efficiency, meaning the engine or powertrain is able to convert more of the fuel energy into useful work with minimal losses. Fig. 11 demonstrate the combined effect of CR, fuel blend, injection pressure and CeO_2 nanoparticles on BSEC while keeping the other values at hold. Fig. 11(a) illustrate the effect of fuel blends and CR on BSEC. There is a marginal change in BSEC concerning to CR, it was observed as the fuel blends increases the BSEC reduces to 13.49 MJ/kW-h at 10.4 % V/V and it increases further for higher values as the blends increases to 15 % V/V. Sayin et al. [55] also observed a similar pattern of increase in BSEC with fuel blend. Fig. 11(b) demonstrate that optimized value of BSEC is obtained at the range of nanoparticle concentration in between 90.23 ppm to 118.15 ppm while having the fuel blends concentration in range of 6.67 to 11.58 % v/v. Fig. 11(c) illustrate the minimum value of BSEC of 13.5 MJ/kW-h in the fuel blends range of 6.58 to 12.01 v/v with the injection pressure of 791.94 bar. Fig. 11(d) depicts the variation of BSEC

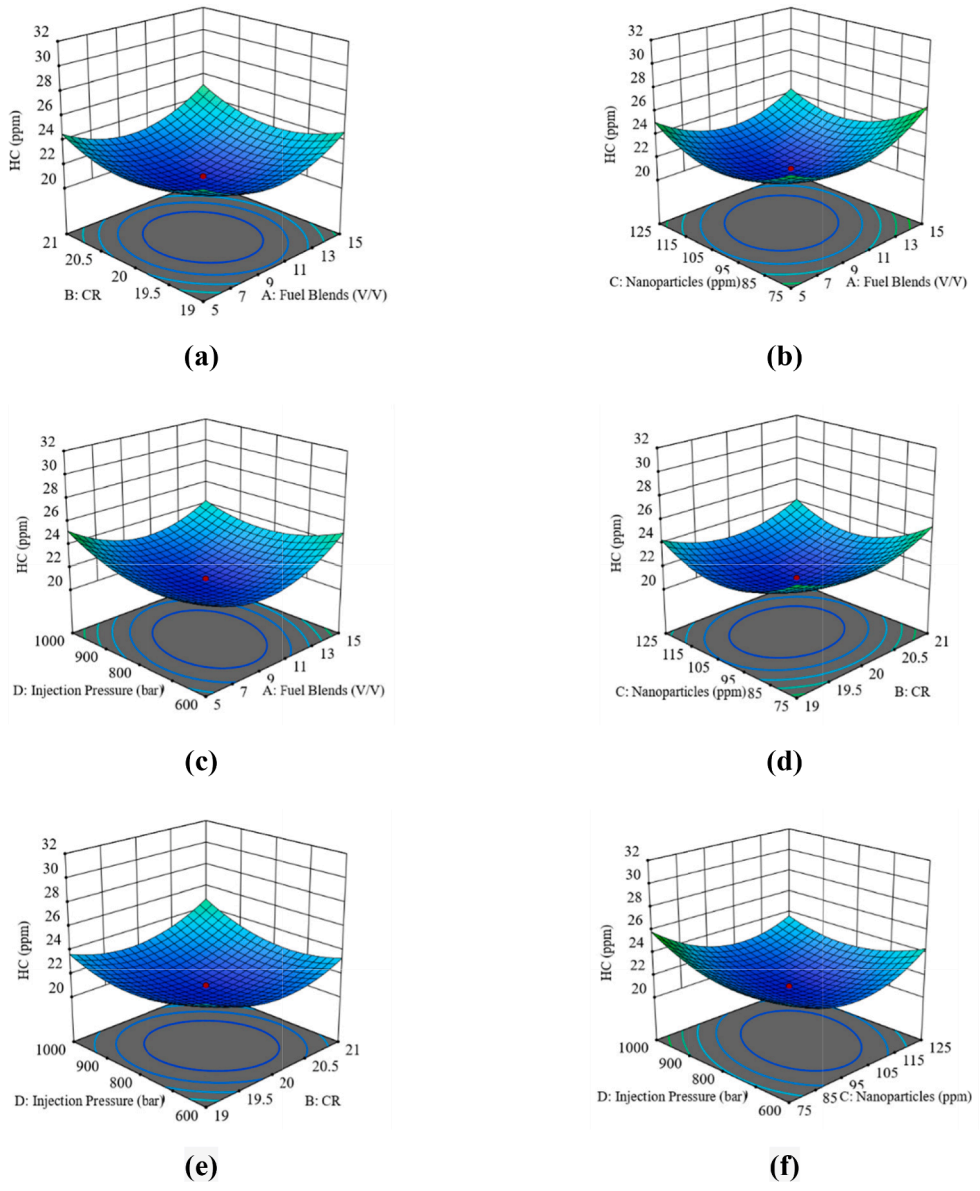


Fig. 14. Variation of HC emission against CR, fuel blend, injection pressure, and CeO_2 nanoparticles.

with nanoparticles range and compression ratio. The optimum value were found in the range of nanoparticles in 90.6 to115.4 ppm, whereas a minimal variation of BSEC were found of 13.6 MJ/kW-h at CR of 19 to 13.4 MJ/kW-h at CR 21. Fig. 11(e) shows the variation of injction pressure of 734.67 bar to 923.86 for the optimum value of BSEC for the range of minimalist compression ratio. The range of injection pressure with variation in nanoparticle is shown in Fig. 11(f). It depicts the optimum value of BSEC is found at the injection pressure of 102.9 bar with the concentration of nanoparticles of 802.14 ppm.

4.1.2. Effect of control factors on brake thermal efficiency

The efficiency with which an engine transfers the energy content of fuel into usable work is referred to as brake thermal efficiency (BTE). The influence of input parameters on BTE with hold values is depicted in Fig. 12. Fig. 12(a) shows the cumulative effect of fuel blend % and CR on BTE. It was found that there is a slight variation in BTE due to the fuel blend % since it is 9.9 % in actual concentration. The biodiesel has a higher CV than butanol, therefore it is marginally increased with blend %. Furthermore, BTE improves with increasing CR up to 20.7, after which it begins to decline. This might be because, with greater compression ratio, the engine may face greater heat losses and pumping losses, resulting in a lower BTE. When changing the load to the different fuel blend ratios, O D Onukwuli et al. [56] discovered an analogous set of findings. Fig. 12(b) depicts the interaction effect of CeO_2 nanoparticles and fuel blend on BTE. The addition of nanoparticles to the test fuel raises the BTE, as illustrated. Because of their huge surface area and catalytic impact, nanoparticles increase BTE, resulting in more fuel oxidizer contact and rate of reaction, which enhances combustion efficiency. The optimum value of BTE is observed within the range of nanoparticles in 85.4 to 114.9 with the variation of fuel blends of 7.17 to 12.8 % v/v. Fig. 12(c) depicts the variation of injection pressure and fuel blend on BTE, it shows optimum range of injection pressure within 688.5 bar to 947.7 bar for range of fuel blends within 7.07 to 12.5 % v/v. Fig. 12(d) illustrate that optimum value of BTE is observed at CR higher than 19.6 for the range of nanoparticles in 86.2 to 113.34 ppm. Furthermore, higher BTE is associated with higher compression ratios. A higher CR raises the temperature and pressure of the air in the combustion chamber, enabling better fuel atomizing and blending. This results in more consistent and efficient combustion, which improves BTE [57]. Fig. 12(e) depicts the impact of CR and injection pressre on the BTE

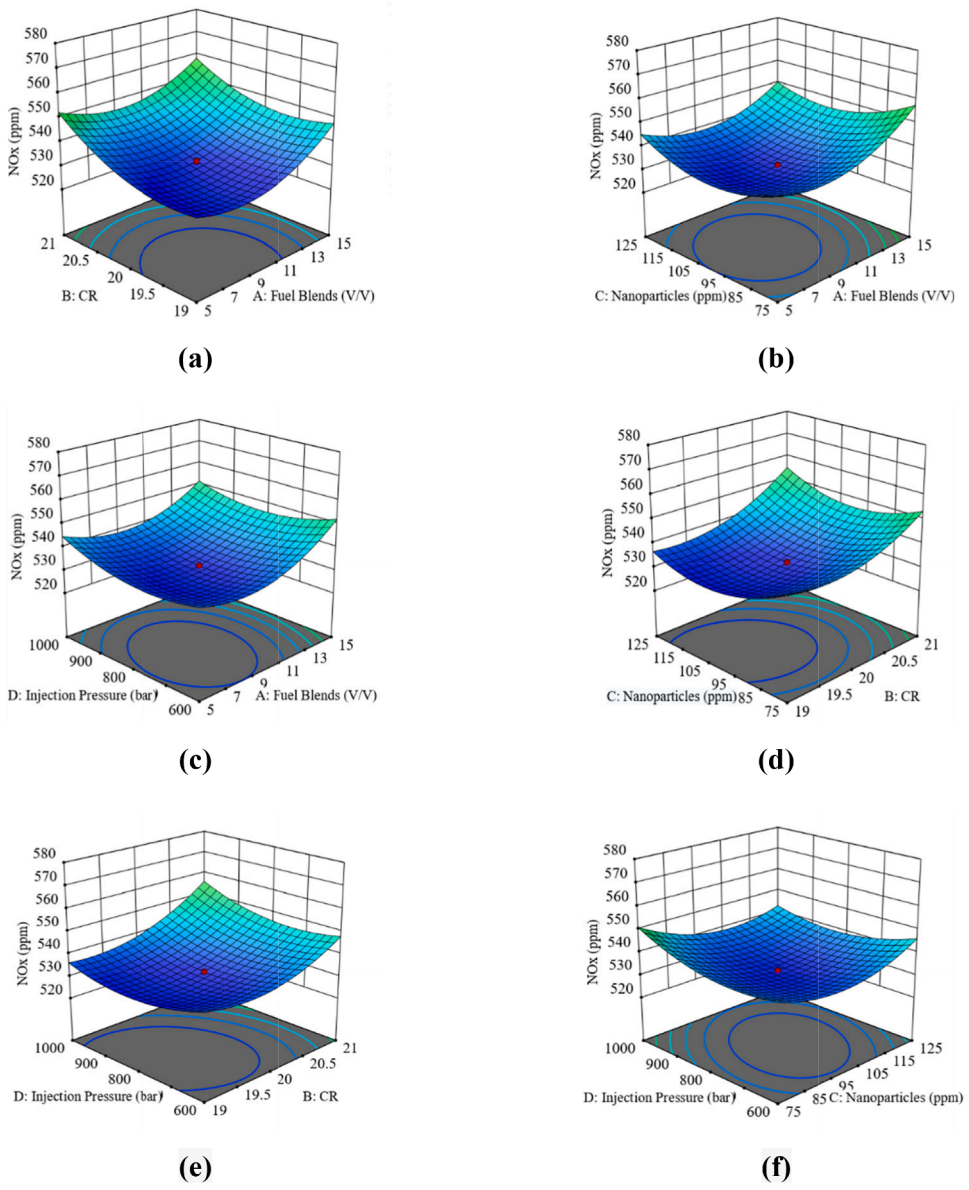


Fig. 15. Variation of NOx emission against CR, fuel blend, injection pressure, and CeO_2 nanoparticles.

of the CRDI engine, which showed that CR above 19.6 improves the BTE since higher CR allows for more efficient combustion of the air–fuel mixture. The range of injection pressure is observed within 682.9 to 923.9 bar. The effect of CeO_2 NP and the injection presure was reported in Fig. 12(f), there is a drastic increase in BTE due to nanoparticle since the engine can operate closer to its optimal operating point, resulting in higher BTE. For the nanoparticle variation of 83.5 to 111.34 for injection pressure variation in between 693.8 bar to 935.47 bar optimized value of BTE is obtained.

4.1.3. Effect of controlling parameters on carbon monoxide

The main factors influencing the generation of carbon monoxide (CO) emissions include a lack of sufficient oxygen for complete combustion, inadequate mixing of air and fuel, areas with higher fuel concentration, and moderate combustion temperatures. Fig. 13 depicts the influence of input parameters on CO, alongside the hold values. Fig. 13 (a) demonstrates that when the blend ratio is increased, the response CO decreases and the load increases until it reaches a minimum at blend 10.16 % for a CR of 20.04. This might result from the combustion process being less effective at a high compression ratio, increasing CO emissions. Fig. 13(b) displays the effect of CeO_2 NP and fuel blend on CO emission while maintaining other parameters at hold. It was discovered that the inclusion of CeO_2 nanoparticles reduced CO emissions. The minimum value of CO emission is found at a concentration of 98.17 ppm of nanoparticles. This can be attributed to enhanced fuel–air blending, accelerated combustion due to catalytic activity, and reduced ignition delays, all of which contribute to more efficient combustion and lower CO emissions. Fig. 13(c) shows the range of 600 to 875.44 for the variation of injection pressure for obtaining the optimised value of CO for the fuel blend in variation of 7.45 to 13.6 % v/v. Fig. 13(d) depicts the variation of nanoparticles and compression ratio on the CO. Minimized value of CO is observed for the range of CR in 19.08 to 20.44 for concentration of nanoparticles in range of 82.6 to 113.9 ppm. This is because compressing the air–fuel combination to a smaller volume before ignition promotes greater atomization and mixing of the fuel with air. Also, a higher compression ratio leads to a lesser amount of CO emission as shown in the surface plot. This can result in more complete combustion, minimising the amount of unburned fuel and, as a result, CO emissions. Earlier studies by C Srinidhi et al. [58] had similar findings. Fig. 13(e) and 13(f) showed the cumulative effect of injection

pressure and compression ratio, injection pressure and nanoparticle, respectively, on CO emission. The CO emission becomes lesser by enhancing the NP concentration to 101.9 ppm and CR range of 19.1 to 20.5. The air–fuel ratio inside the combustion chamber is affected by the CR. A greater CR can cause the air–fuel mixture to have a higher air-to-fuel ratio. Leaner mixes with more air and less fuel tend to emit less CO.

4.1.4. Effectof control factors on hydrocarbon

Fig. 14 depicts the surface plots of the Hc emission. There is a marginal change in Fig. 14(a) owing to fluctuation in the fuel blend ratio since it is 10.13 % in concentration. However, the quantity of HC emissions produced by a CRDI diesel engine rises as the compression ratio increases to 20.5. Higher compression ratio need more fuel to be fed into the combustion chamber, which can result in incomplete combustion and the generation of HC emissions. Fig. 14(b) demonstrated the iterative effect of CeO_2 NP and fuel blend on the emission of HC while holding the other value. It shows that minimum value of 21 ppm HC is obtained at 99.4 ppm concentarion of nanoparticles and 9.95 % of fuel blend having the optimized range of 86.8 to 116.4 ppm and 6.9 to 12.6 % v/v of CeO_2 NP concentration and fuel blend respectively. Fig. 14(c) shows the varitation of injection pressure and fuel blend ratio on HC emissions the optimal range of injection pressure is observed in the range of 642.8 bar to 906.24 bar for the range of blend percentage in 6.8 % to 12.67 %. Considering biodiesel's substantial amount of oxygen minimises nearby affluent mixture regions and a higher compression ratio lowers the temperature where it fires up, it should emit fewer hydrocarbons. Increasing the CR reduces hydrocarbon emissions, as seen in Fig. 14(d). The concentration of CeO_2 nanoparticles increases to more than 117.3 ppm, the emission of HC also increases. This is because nanoparticles have a higher catalytic efficiency and a significant surface-to-volume ratio, which improves fuel fragmentation and burning, resulting in reduced levels of hydrocarbon (HC) emissions.The acquired results were more consistent with those of Gavhane et al. [59]. The cumulative effect of injection pressure and compression ratio, injection pressure and nanoparticle depicted in Fig. 14 (e) and (f) on the variation in the emission of HC. CeO_2 nanoparticles are thought to have catalytic capabilities that can improve combustion efficiency and minimise HC emissions. These nanoparticles may increase fuel atomization, air–fuel mixing, and combustion efficiency, resulting in fewer unburned hydrocarbons in the exhaust.

4.1.5. Effect of control factors on NOx

Fig. 15 depicts the combined impacts of CR, fuel blend ratio, injection pressure, and CeO_2 nanoparticles concentration on NOx emission with hold values. Fig. 15(a) showed the effect of fuel blend % and CR on the variation of NOx emission while the hold values were CeO_2 NP 100 ppm and Injection pressure at 800 bar. Minimised value of NOx were found for the blend ratio of less than 11.8 % v/v and the CR below 20.07. There is a slight variation in the NOx emission due to a change in the fuel blend ratio. It was also discovered that NOx emissions rise with compression ratio. This is because heavier CR necessitates more fuel burning, resulting in higher temperatures and more nitrogen oxide (NOx) generation. At higher CR, the combustion process is more vigorous, which might encourage the production of NOx particles. Fig. 15(b) showed that raising the NP concentration with fuel blend causes an increase in NOx production. The effect of nanoparticles on NOx emissions varies from 81.15 ppm to 117.7 ppm, depending on the nanoparticles utilised and their characteristics. In some situations, nanoparticles can operate as catalysts, promoting more complete combustion and lowering NOx emissions. Fig. 15(c) depicts the variation of injection pressure and fuel blend % on the emission of NOx and its show the optimum value in the range of injection pressure in between 611.41 bar to 916.34 bar for the range of fuel blend in 5.3 % to 11.4 % v/v. Fig. 15(d) illustrate the variation of NOx for the range of nanoparticle concentration in between 87.3 ppm to 120.8 ppm, whereas keeping the CR below 20.2. In internal combustion engines, higher compression ratios tend to increase NOx emissions. This is owing to the increased temperatures and pressures encountered during the combustion process. Higher peak temperatures arise from increased compression ratios, which can aid in the generation of thermal NOx [60]. The cumulative effect of of injection pressure and compression ratio, injection pressure and nanoparticle depicted in Fig. 15(e) and (f) on the variation in the emission NOx. As the compression ratio rises, the NOx emission value rises as well. The inclusion of additional oxygen particles in biodiesel raises the oxygen level within the combustion chamber, boosting NOx output.

4.2. Comparative analysis of experimental and predicted value of engine emissions and performance responses

The comparative analysis of engine emission and performance is depicted in Figs. 16 to 20, utilising RSM and MLP regressor models.

Fig. 16 depicts the variations in the expected and observed values of BSEC. It is worth noting that both the RSM and MLP regressor models show a drop in performance, particularly in the experimental data range of 5–10. Interestingly, when analysing the coefficient of determination (R^2), the MLP regressor models have the greatest R^2 value as compared

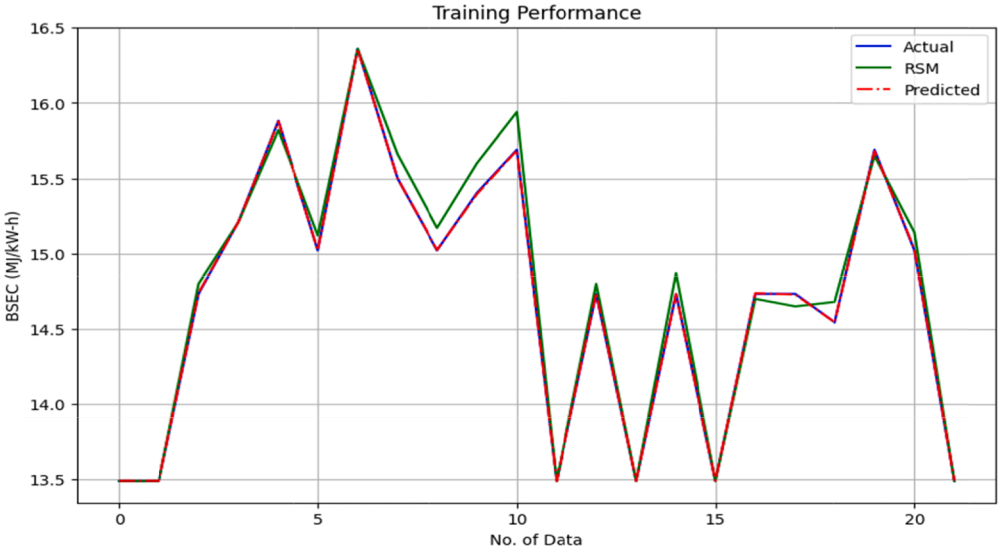


Fig. 16. Comparison of experimental, RSM, and MLP regressor models for BSEC.

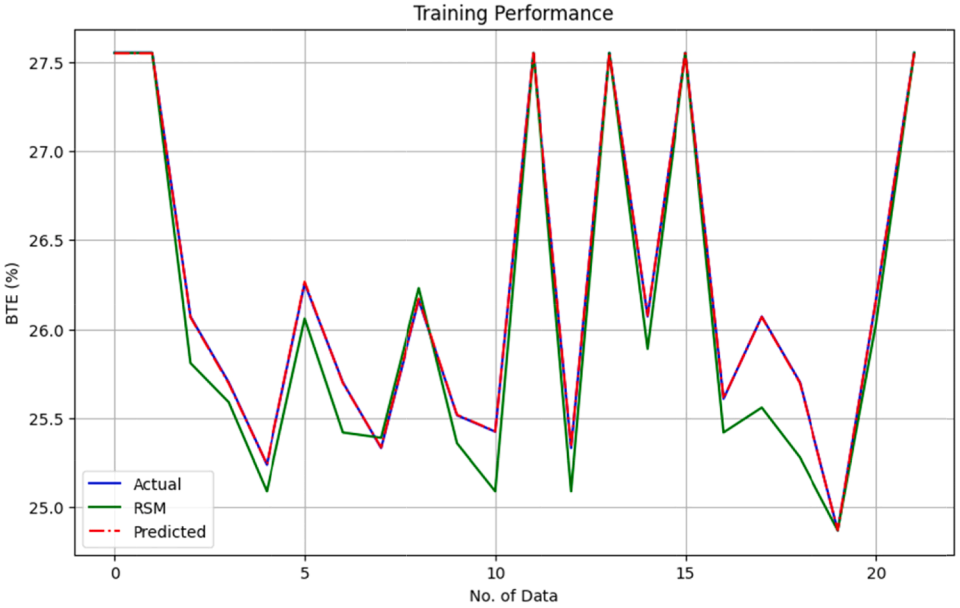


Fig. 17. Comparison of experimental, RSM, and MLP regressor models for BTE.

to RSM model. This gap could be traced to the RSM model's poor performance, as evidenced by the 5th and 10th sets of experimental data.

Fig. 17 compares the BTE forecasting results to the RSM and MLP regressor models, offering a comparison of their performance. Table 12 elaborates on these findings, showing that the R^2 values ranged between 0.9967 and 0.9998 across the models. Notably, the MLP regressor model has the greatest R^2value, at an amazing 0.9998. According to evaluation standards, prediction outcomes are considered "excellent" when the RMSE value is low or close to zero [42,44]. According to the extant research, both models produced impressive outcomes. Furthermore, the MLP regressor-based model distinguishes itself by producing "outstanding" results, highlighting its superiority in forecasting accuracy.

Fig. 18 shows the differences between the projected CO values of the RSM and MLP regressor models and the actual experimental results. A careful examination of Fig. 17 reveals the effectiveness of both predictive models in anticipating CO levels. The findings given in Table 12 provide more evidence of its effectiveness. Notably, in terms of CO prediction, the MLP regressor model has the highest R^2 value, indicating higher performance. Nonetheless, the RSM model's R^2 values are very similar to those of the MLP regressor model, demonstrating its comparative performance.

Fig. 19 shows the differences in projected HC values between the RSM and MLP regressor models as compared to experimental values. Fig. 16 shows that the RSM model performed poorly when predicting HC. Notably, its predictions differed significantly from those produced by the MLP regressor model, particularly for data ranging from 2 to 10. This deficiency is further supported by statistical analyses, which highlight the RSM model's shortcomings in this area. Furthermore, the mismatch underlines the need for improvement or other methodologies in predictive modelling to achieve correct HC estimates.

Fig. 20 compares the NOx values projected by the RSM and MLP regressor models to the experimental data. Notably, the predictive models' NOx predictions are quite accurate, as shown in the image. Table 12 contains further information about the performance of these models, including significant statistical measurements. The R^2 value for

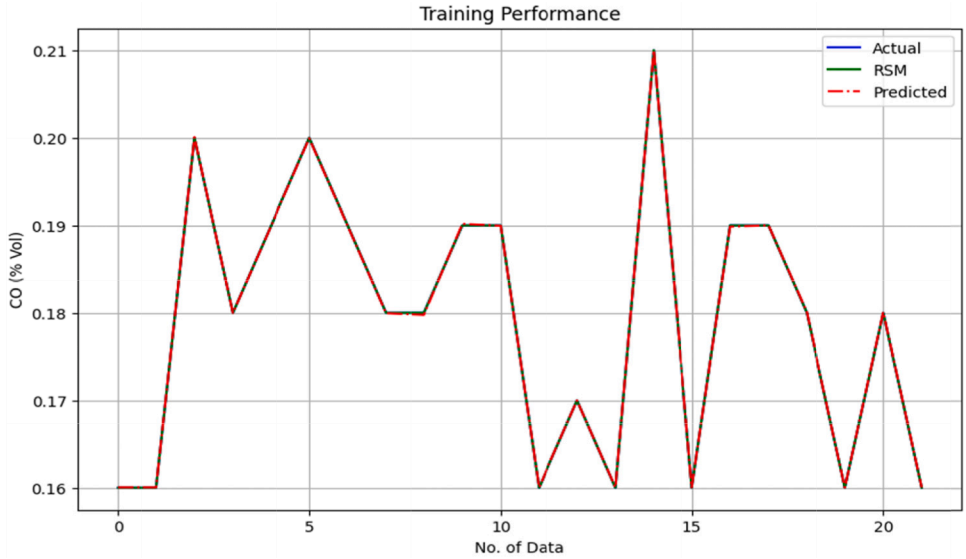


Fig. 18. Comparison of experimental, RSM, and MLP regressor models for CO(%).

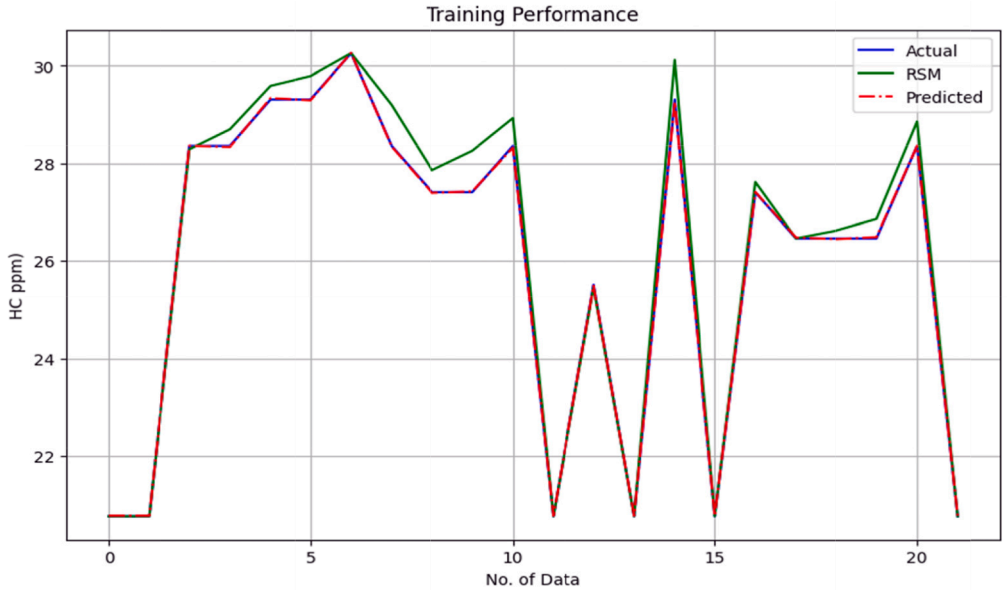


Fig. 19. Comparison of experimental, RSM, and MLP regressor models for HC.

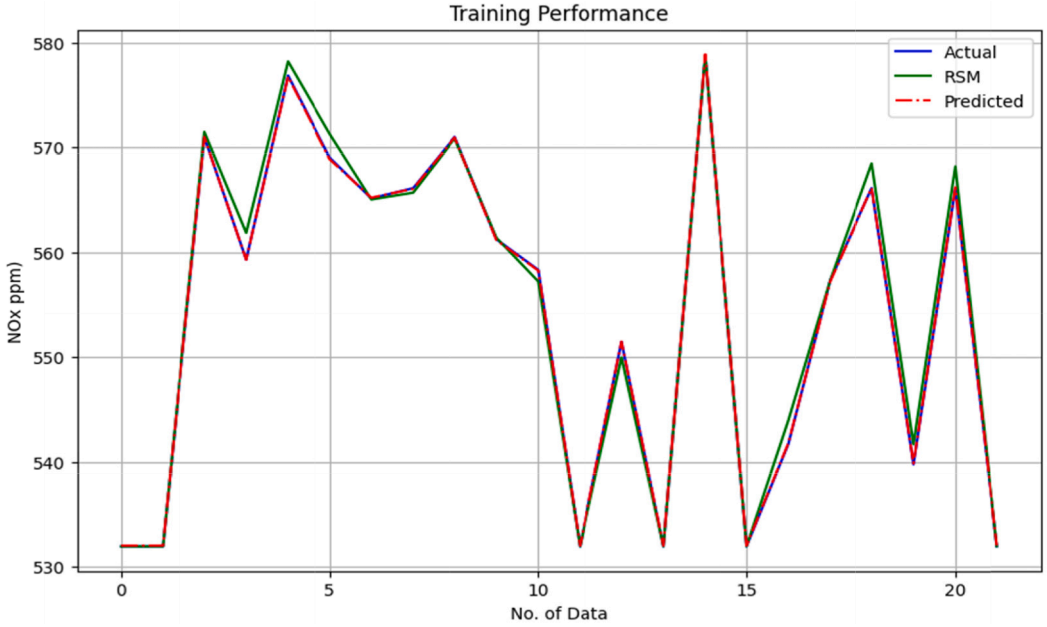


Fig. 20. Comparison of experimental, RSM, and MLP regressor models for NOx.

NOx prediction using the MLP regressor was an astounding 0.9998, above that of the RSM model at 0.8998. However, it is worth noting that, despite having a higher R^2 value, the MLP regressor model produced much worse predictions than the RSM model, as seen in observations 3, 4, and 5. This difference is further explained by statistical analysis.

The comparative analysis of engine emission and performance is depicted in Figs. 16 to 20, utilising RSM and MLP regressor models. The graphical representations clearly demonstrate that the MLP regressor model consistently exhibits superior accuracy in predicting values compared to the RSM-based models. This observable pattern leads to the

Table 11
Validation and experimental confirmation tests for input factor.

Response	Experimental responses on optimised process parameters	Predicted responses by RSM	Predicted responses by MLP	% Error with respect to RSM	% Error with respect to MLP
BSEC (MJ/kW-h)	13.9	13.5	13.47	2.87	3.09
BTE (%)	28.7	27.8	27.55	3.13	4.00
CO (%vol.)	0.17	0.16	0.16	3.85	3.85
HC (ppm)	21.44	20.996	20.685	2.14	3.52
NOx (ppm)	551.77	531.989	531.823	3.58	3.61

Table 12
Statistical indicators estimated values.

Factors	Statistical Matrices	Predictive models	
		RSM	MLP
BSEC	MAE	0.0815	0.0650
	R^2	0.8987	0.9987
	RMSE	0.0141	0.0015
BTE	MAE	0.0085	0.0640
	R^2	0.9887	0.9998
	RMSE	0.0376	0.0017
CO	MAE	0.0103	0.0598
	R^2	0.8987	0.9967
	RMSE	0.0444	0.0016
HC	MAE	0.0157	0.0681
	R^2	0.8897	0.9977
	RMSE	0.0560	0.0023
NOx	MAE	0.0193	0.0716
	R^2	0.8998	0.9998
	RMSE	0.1365	0.0167

conclusion that the MLP regressor model surpasses the RSM-based models in accurately predicting engine emission and performance. Consequently, this finding emphasises the effectiveness of the MLP regressor approach in achieving more precise predictions within this specific context.

4.3. Confirmation of the result

To assess the correctness of the findings achieved by optimisation and MLP regressor, a validation test was done in combination with optimisation; for verification, a test was performed using the optimal engine settings obtained during optimisation. Due to experimental limitation, optimised process parameters are considered in the absolute fixation such as 9.97 % fuel blend is taken as 10 %, injection pressure is taken as 800 bar, and CR is taken as 20. Table 11 displays the results of the optimal and test validations and the degree of error across them. The recorded discrepancy is approximately 5 %, which is deemed within an acceptable range.

4.4. Phase of prophecy

Table 12 summarises the numerical data derived from the statistical indicators employed to all of the study's anticipated data. Table 10 illustrates that the MLP regressor's RMSE values generate good results. Furthermore, the MAE values of the forecasts are typically close to zero. As a result, the algorithm-generated parameter averages are typically lower than the actual observed data averages. Notably, when the RMSE values are compared, the MLP regressor predictions exhibit lower RMSE values than the RSM-based models. When compared to RSM models, these findings show the efficacy of MLP regressors in providing accurate predictions with higher statistical performance.

5. Conclusions

The results of this study showed the possibility of combining the blend ratio of Karanja biodiesel and diesel with 10 % Biobutanol, CR, IP, and CeO_2 nanoparticle concentration in a CRDI engine were explored in the present investigation. The following finding were obtained from the current study.

- The optimal process parameters are achieved a biodiesel blend ratio of 9.96 %, CR of 20, IP 797.4 bar, and a CeO_2 nanoparticle concentration of 100.3 ppm through CCFD model of RSM.
- After optimization the responses were BSEC of 13.49 MJ/kW-h, BTE of 27.8 %, and emitted 0.16 % CO, 20.9 ppm HC, and 531.9 ppm NOx.
- The high R^2 values for each response variable (99.98 % for BTE, 99.97 % for BSEC, 99.67 % for CO, 99.77 % for HC, and 99.98 % for NOx) suggest a significant connection between the input parameters and the response variables.
- The MAE value of the MLP regressor-based model ranged from 0.0011 to 0.0119. In terms of RMSE, MLP regressor-based model outputs are within reasonable boundaries.
- The most detrimental RMSE findings were reported in NOx anticipate for both. Only RSM-based models exceed 10 % in NOx emissions.
- The accuracy of the results obtained by optimisation and MLP regression; a validation test was conducted to confirm; the reported disparity is around 5 %, which is considered within acceptable limits.

This study suggests that sustainable and efficient engine technology can be achieved by methodically investigating parameter space and determining optimal operating conditions. It also highlights the potential benefits of using a combination of Karanja biodiesel and Biobutanol with diesel, along with CeO_2 nanoparticles, to improve the performance and emissions characteristics of CRDI engines.

Declarations

Ethical approval

This study does not require ethical approval because it is not conducted on humans or animals.

Funding

There is no financial support provided for this work.

CRediT authorship contribution statement

Ashish Kumar Singh: Writing – original draft, Validation, Methodology, Conceptualization. **Harveer Singh Pali:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization. **Nishant K. Singh:** Writing – review & editing, Resources, Methodology, Conceptualization. **Sanjeev Kumar:** Writing – review & editing, Visualization, Methodology, Data curation. **Abhishek Sharma:** Writing – review & editing, Validation, Resources, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] Rajpoot AS, Choudhary T, Chelladurai H, Nath Verma T, Pugazhendhi A. "Sustainability analysis of spirulina biodiesel and their blends on a diesel engine with energy, exergy and emission (3E's) parameters". *Fuel* 2023;349:128637. https://doi.org/10.1016/J.FUEL.2023.128637.
[2] Zhai Y, Li J, He D, Liang H, Sha Y, Li G. Modeling and analysis of a bias-free hydrogen production approach using perovskite photocathode and lignocellulosic biomass. Energy Convers Manag 2023;vol. 298(July):117807. https://doi.org/10.1016/j.enconman.2023117807.
[3] Badawy T, et al. Selection of second-generation crop for biodiesel extraction and testing its impact with nano additives on diesel engine performance and emissions. Energy 2021;237:121605. https://doi.org/10.1016/J.ENERGY.2021.121605.
[4] Sajjad MOA, Sathish T, Saravanan R, Asif M, Linul E, Ağbulut Ü. Combustion, performance and emission discussion of soapberry seed oil methyl ester blends and exhaust gas recirculation in common rail direct fuel injection system. Energy 2023; 278:127763. https://doi.org/10.1016/J.ENERGY.2023.127763.
[5] Nour M, Attia AMA, Nada SA. Combustion, performance and emission analysis of diesel engine fuelled by higher alcohols (butanol, octanol and heptanol)/diesel

blends. Energy Convers Manag 2019;185:313–29. https://doi.org/10.1016/j.enconman.2019.01.105.
[6] Sayyed S, Das RK, Kulkarni K. Performance assessment of multiple biodiesel blended diesel engine and NOx modeling using ANN. Case Stud Therm Eng 2021;28:101509. https://doi.org/10.1016/J.CSITE.2021.101509.
[7] Hosseinzadeh-Bandbafha H, et al. Exergetic, economic, and environmental life cycle assessment analyses of a heavy-duty tractor diesel engine fueled with diesel–biodiesel-bioethanol blends. Energy Convers Manag 2021;241:114300. https://doi.org/10.1016/J.ENCONMAN.2021.114300.
[8] Fayad MA, Alani WK, Dhahad HA, Zheng J. "Diminution of air pollution from NOX and smoke/soot emitted from alcohols/diesel blends in diesel engine and influence of the exhaust gas recirculation (EGR)". *J Environ Eng Landsc Manag* 2023;31(1):103–12. https://doi.org/10.3846/JEELM.2023.17410.
[9] Uyumaz A, et al. Experimental investigation on the combustion, performance and exhaust emission characteristics of poppy oil biodiesel-diesel dual fuel combustion in a CI engine. Fuel 2020;280:118588. https://doi.org/10.1016/J.FUEL.2020.118588.
[10] Nanthagopal K, et al. Study on decanol and Calophyllum Inophyllum biodiesel as ternary blends in CI engine. Fuel 2019;239:862–73. https://doi.org/10.1016/j.fuel.2018.11.037.
[11] Chandran D. Compatibility of diesel engine materials with biodiesel fuel. Renew Energy Mar. 2020;147:89–99. https://doi.org/10.1016/J.RENENE.2019.08.040.
[12] Verma TN, et al. "A comprehensive review of the influence of physicochemical properties of biodiesel on combustion characteristics, engine performance and emissions". *J Traffic Transp Eng (English Ed* 2021;8(4):510–33. https://doi.org/10.1016/J.JTTE.2021.04.006.
[13] Singh D, et al. A comprehensive review of physicochemical properties, production process, performance and emissions characteristics of 2nd generation biodiesel feedstock: Jatropha curcas. Fuel 2021;285:119110. https://doi.org/10.1016/J.FUEL.2020.119110.
[14] Hosseinzadeh-Bandbafha H, Tabatabaei M, Aghbashlo M, Khanali M, Demirbas A. A comprehensive review on the environmental impacts of diesel/biodiesel additives. Energy Convers Manag 2018;174:579–614. https://doi.org/10.1016/J.ENCONMAN.2018.08.050.
[15] Cihan Ö. Experimental and numerical investigation of the effect of fig seed oil methyl ester biodiesel blends on combustion characteristics and performance in a diesel engine. Energy Rep 2021;7:5846–56. https://doi.org/10.1016/J.EGYR.2021.08.180.
[16] Kim HY, Ge JC, Choi NJ. "Effects of fuel injection pressure on combustion and emission characteristics under low speed conditions in a diesel engine fueled with palm oil biodiesel". *Energies* 2019;12(17):3264. https://doi.org/10.3390/EN12173264.
[17] Viswanathan K, Ashok B, Pugazhendhi A. Comprehensive study of engine characteristics of novel biodiesel from curry leaf (Murraya koenigii) oil in ceramic layered diesel engine. Fuel 2020;280:118586. https://doi.org/10.1016/J.FUEL.2020.118586.
[18] Rajak U, Nashine P, Dasore A, Verma TN. Utilization of renewable and sustainable microalgae biodiesel for reducing the engine emissions in a diesel engine. Fuel 2022;311:122498. https://doi.org/10.1016/J.FUEL.2021.122498.
[19] Pourhoseini SH, Ghodrat M, Baghban M, Shams Z. Effects of blending energetic iron nanoparticles in B20 fuel on lower CO and UHC emissions of the diesel engine in cold start condition. Case Stud Therm Eng 2023;41:102658. https://doi.org/10.1016/J.CSITE.2022.102658.
[20] EL-Seesy AI, Xuan T, He Z, Hassan H. Enhancement the combustion aspects of a CI engine working with Jatropha biodiesel/decanol/propanol ternary combinations. Energy Convers Manag 2020;226:113524. https://doi.org/10.1016/j.enconman.2020.113524.
[21] Hussain Vali R, Marouf Wani M. Optimal utilization of ZnO nanoparticles blended diesel-water emulsion by varying compression ratio of a VCR diesel engine. J Environ Chem Eng 2020;vol. 8(4). https://doi.org/10.1016/J.JECE.2020.103884.
[22] Mohammed AS, Atnaw SM, Ramaya AV, Alemayehu G. A comprehensive review on the effect of ethers, antioxidants, and cetane improver additives on biodiesel-diesel blend in CI engine performance and emission characteristics. J Energy Inst 2023;vol. 108(March):101227. https://doi.org/10.1016/j.joei.2023.101227.
[23] Pullagura G, et al. "Influence of dimethyl carbonate and dispersant added graphene nanoplatelets in diesel-biodiesel blends: Combustion, performance, and emission characteristics of diesel engine". *Int J Energy Res* 2023;2023(Ci). https://doi.org/10.1155/2023/9989986.
[24] Chacko N, Jeyaseelan T. Comparative evaluation of graphene oxide and graphene nanoplatelets as fuel additives on the combustion and emission characteristics of a diesel engine fuelled with diesel and biodiesel blend. Fuel Process Technol 2020;204:106406. https://doi.org/10.1016/J.FUPROC.2020.106406.
[25] Anbarsooz M. Combustion characteristics of nanofuels: A comprehensive review on diesel/biodiesel-based nanofuels. Fuel 2023;337:126834. https://doi.org/10.1016/J.FUEL.2022.126834.
[26] Reddy SNK, Wani MM. A comprehensive review on effects of nanoparticles-antioxidant additives-biodiesel blends on performance and emissions of diesel engine. Appl Sci Eng Prog 2020;vol. 13(4). https://doi.org/10.14416/J.ASEP.2020.06.002.
[27] Tomar M, Kumar N. "Influence of nanoadditives on the performance and emission characteristics of a CI engine fuelled with diesel, biodiesel, and blends – a review". Energy Sources Part A Recover Util Environ Eff 2020;42(23):2944–61. https://doi.org/10.1080/15567036.2019.1623347.
[28] Yusuf AA, et al. Effect of low CeO2 nanoparticles dosage in biodiesel-blends on combustion parameters and toxic pollutants from common-rail diesel engine. Atmos Pollut Res 2022;13(2):101305. https://doi.org/10.1016/J.APR.2021.101305.
[29] Ge S, et al. Performance, combustion and emission characteristics of the CI engine fueled with botryococcus braunii microalgae with addition of TiO2 nanoparticle. Fuel 2022;317:121898. https://doi.org/10.1016/J.FUEL.2021.121898.
[30] Dinesha P, Kumar S, Rosen MA. Effects of particle size of cerium oxide nanoparticles on the combustion behavior and exhaust emissions of a diesel engine powered by biodiesel/diesel blend. Biofuel Res J 2021;8(2):1374–83. https://doi.org/10.18331/BRJ2021.8.2.3.
[31] Singh AK, Pali HS, Khan MM. Response surface optimization approach to enhance the performance of diesel engine using ternary blends of biobutanol and biodiesel doped with ZnO nanoparticles. Energy Convers Manag 2024;300:117899. https://doi.org/10.1016/J.ENCONMAN.2023.117899.
[32] Ge JC, Wu G, Yoo BO, Choi NJ. Effect of injection timing on combustion, emission and particle morphology of an old diesel engine fueled with ternary blends at low idling operations. Energy 2022;253:124150. https://doi.org/10.1016/j.energy.2022.124150.
[33] Singh Pali H, Sharma A, Kumar N, Singh Y. Biodiesel yield and properties optimization from Kusum oil by RSM. Fuel 2021;291:120218. https://doi.org/10.1016/J.FUEL.2021.120218.
[34] Yan J, Gao S, Zhao W, Lee TH. Study of combustion and emission characteristics of a diesel engine fueled with diesel, butanol-diesel and hexanol-diesel mixtures under low intake pressure conditions. Energy Convers Manag 2021;243:114273. https://doi.org/10.1016/j.enconman.2021.114273.
[35] Kumar P, Kumar N. "Study of ignition delay period of n-Butanol blends with JOME and diesel under static loading conditions". Energy Sources Part A Recover Util Environ Eff 2018;40(14):1729–36. https://doi.org/10.1080/15567036.2018.1486904.
[36] Praveen A, Lakshmi Narayana Rao G, Balakrishna B. Performance and emission characteristics of a diesel engine using Calophyllum Inophyllum biodiesel blends with TiO2 nanoadditives and EGR. Egypt J Pet 2018;27(4):731–8. https://doi.org/10.1016/j.ejpe.2017.10.008.
[37] Cai Y, Jia M, Li Y, Cao J, García A, Monsalve-Serrano J. Multiple combustion modes switching to realize full-load efficient energy conversion of n-butanol/diesel dual direct injection (DI2) engine. Energy Convers Manag 2023;vol. 278 (February):116722. https://doi.org/10.1016/j.enconman.2023.116722.
[38] Subramani S, Govindasamy R, Rao GLN. Predictive correlations for NOx and smoke emission of DI CI engine fuelled with diesel-biodiesel-higher alcohol blends-response surface methodology approach. Fuel 2020;269:117304. https://doi.org/10.1016/J.FUEL.2020.117304.
[39] Ashok B, et al. Multi-objective optimization on vibration and noise characteristics of light duty biofuel powered engine at idling condition using response surface methodology. J Energy Resour Technol Trans ASME 2021;vol. 143(4). https://doi.org/10.1115/1.4047974/1085950.
[40] Vinay BS, Yadav AK. Optimisation of performance and emission characteristics of CI engine fuelled with Mahua oil methyl ester–diesel blend using response surface methodology. Int J Ambient Energy May 2020;41(6):674–85. https://doi.org/10.1080/01430750.2018.1484804.
[41] Mehra KS, Goel V, Kumar R. An integrated multi-attribute decision framework for sustainability assessment of renewable diesel fuel production pathways. Energy Convers Manag 2024;309:118461. https://doi.org/10.1016/J.ENCONMAN.2024.118461.
[42] Tosun E, Ozgur T, Ozgur C, Ozcanli M, Serin H, Aydin K. Comparative analysis of various modelling techniques for emission prediction of diesel engine fueled by diesel fuel with nanoparticle additives. Eur Mech Sci 2017;1(1):15–23. https://doi.org/10.26701/EMS.320490.
[43] Liu J, Huang Q, Ulishney C, Dumitrescu CE. Machine learning assisted prediction of exhaust gas temperature of a heavy-duty natural gas spark ignition engine. Appl Energy 2021;vol. 300(March):117413. https://doi.org/10.1016/j.apenergy.2021.117413.
[44] Rahimi Molkdaragh R, Jafarmadar S, Khalilaria S, Soukht Saraee H. Prediction of the performance and exhaust emissions of a compression ignition engine using a wavelet neural network with a stochastic gradient algorithm. Energy Jan. 2018;142:1128–38. https://doi.org/10.1016/j.energy.2017.09.006.
[45] Liu J, Ulishney C, Dumitrescu CE. Random forest machine learning model for predicting combustion feedback information of a natural gas spark ignition engine. J Energy Resour Technol Trans ASME 2021;143(1):1–7. https://doi.org/10.1115/1.4047761.
[46] Soukht Saraee H, Taghavifar H, Jafarmadar S. Experimental and numerical consideration of the effect of CeO2 nanoparticles on diesel engine performance and exhaust emission with the aid of artificial neural network. Appl Therm Eng 2017;113:663–72. https://doi.org/10.1016/J.APPLTHERMALENG.2016.11.044.
[47] Liu J, Wang H. Machine learning assisted modeling of mixing timescale for LES/PDF of high-Karlovitz turbulent premixed combustion. Combust Flame 2022;238:111895. https://doi.org/10.1016/j.combustflame.2021.111895.
[48] Karagöz M. ANN based prediction of engine performance and exhaust emission responses of a CI engine powered by ternary blends. Int J Automot Sci Technol 2020;4(3):180–4. https://doi.org/10.30939/IJASTECH.771789.
[49] Ağbulut Ü, Sarıdemir S. Is the ethanol additive more environmentally friendly for a spark ignition (SI) engine or for a compression ignition (CI) engine? Environ Eng Manag J 2020;19(4):633–42. https://doi.org/10.30638/eemj.2020.060.
[50] Liu J, Huang Q, Ulishney C, Dumitrescu CE. Comparison of random forest and neural network in modeling the performance and emissions of a natural gas spark ignition engine. J Energy Resour Technol Trans ASME 2022;144(3):1–12. https://doi.org/10.1115/1.4053301.

[51] Liu Z, Liu J. Machine learning assisted analysis of an ammonia engine performance. J Energy Resour Technol Trans ASME 2022;144(11):1–12. https://doi.org/10.1115/1.4054287.
[52] Kumar A, Pali HS, Kumar M. Effective utilization of waste plastic derived fuel in CI engine using multi objective optimization through RSM. Fuel 2024;355(July). https://doi.org/10.1016/j.fuel.2023.129448.
[53] Jacob A, Ashok B. An interdisciplinary review on calibration strategies of engine management system for diverse alternative fuels in IC engine applications. Fuel 2020;vol. 278(March):118236. https://doi.org/10.1016/j.fuel.2020.118236.
[54] Singh Pali H, et al. Enhancement of combustion characteristics of waste cooking oil biodiesel using TiO2 nanofluid blends through RSM. Fuel 2023;331(P1):125681. https://doi.org/10.1016/j.fuel.2022.125681.
[55] Sayin C, Ozsezen AN, Canakci M. The influence of operating parameters on the performance and emissions of a DI diesel engine using methanol-blended-diesel fuel. Fuel 2010;89(7):1407–14. https://doi.org/10.1016/J.FUEL.2009.10.035.
[56] Onukwuli OD, Esonye C, Ofoefule AU. Combustion exhaust release impact, diesel engine performance, and optimization studies of green diesel-petrodiesel blend in a high compression ratio direct-injection compression-ignition engine. Adv Mech Eng 2021;13(5). https://doi.org/10.1177/16878140211018778/ASSET/IMAGES/LARGE/10.1177_16878140211018778-FIG17.JPEG.
[57] Sharma A, Pali HS, Kumar M, Singh NK, Singh Y, Singh D. Study the effect of optimized input parameters on a CRDI diesel engine running with waste frying oil methyl ester-diesel blend fuel with ZnO nanoparticles: a response surface methodology approach. Biomass Convers Biorefinery 2023;13(14):13127–52. https://doi.org/10.1007/S13399-021-02158-6/TABLES/11.
[58] Srinidhi C, Madhusudhan A, Channapattana SV, Gawali SV, Aithal K. RSM based parameter optimization of CI engine fuelled with nickel oxide dosed Azadirachta indica methyl ester. Energy Nov. 2021;234:121282. https://doi.org/10.1016/J.ENERGY.2021.121282.
[59] Gavhane RS, et al. "Effect of zinc oxide nano-additives and soybean biodiesel at varying loads and compression ratios on VCR diesel engine characteristics". *Symmetry 2020* 2020;12(6):1042. https://doi.org/10.3390/SYM12061042.
[60] Zhang Z, et al. Performance, combustion and emission characteristics investigations on a diesel engine fueled with diesel/ ethanol /n-butanol blends. Energy 2022;249:123733. https://doi.org/10.1016/J.ENERGY.2022.123733.

Welding in the World
https://doi.org/10.1007/s40194-024-01805-6

RESEARCH PAPER

Evaluation of the role of process parameters of CMT welding and their statistical behaviour on the tensile properties of weld joint of novel aluminium hybrid composite

Vibhu Singh[1] · **M. S. Niranjan**[1] · **Qasim Murtaza**[1]

Received: 29 April 2024 / Accepted: 26 June 2024
© International Institute of Welding 2024

Abstract
The present study investigates the cold metal transfer (CMT) welding of a novel hybrid aluminium composite (AA6061-T6/ 2wt.% TiB_2/ 0.5wt.% La_2O_3) fabricated via stir casting. It analyzes the effect of welding parameters, i.e. current, travel speed, and gas flow rate, on the tensile strength of the composite's weld joint using response surface methodology (RSM). Also, a dedicated study of the microstructure and mechanical properties of weld joints at optimal parameters has been conducted. Results indicate that current significantly affects tensile strength, followed by travel speed and gas flow rate. The optimal parameters identified are 142 A current, 9 mm/s travel speed, and 14 L/min gas flow rate for maximizing tensile strength. Welds at optimal parameters displayed no solidification cracking and minimum porosity. The microstructural analysis confirmed the presence of reinforcements in the composite with the formation of finer grains in the fusion zone (FZ) compared to the heat-affected zone (HAZ). Microhardness tests showed the highest values in the FZ and the lowest in the HAZ, while tensile tests showed reduced strength at HAZ compared to the FZ and base composite, with dominant brittle behaviour. Post-weld heat treatment improved ductility, as indicated by deeper dimples and tear ridges with fewer cleavage facets in fracture analysis.

Keywords MMCs · CMT · Response surface methodology · Rare earth oxide · Tensile strength · Microstructure · Microhardness

1 Introduction

In the current context of globalization, the demand for materials with advanced properties and the capability to be used in state-of-the-art process technologies is increasing at an exponential rate. Metal matrix composite (MMC) can satisfy these needs as it is an engineered material known to have metal as the matrix and one or more reinforcements with different forms like particles, fibres, and whiskers. MMC has replaced a large number of traditional monolithic materials from various industries like aerospace, automobile, and defence as monolithic materials do not offer the advantage of tailoring the properties of materials as MMC has the capacity of displaying excellent properties like high strength, lesser weight, enhanced fatigue properties, better modulus of elasticity, and good wear resistance. These properties can be controlled by the proper choice of reinforcement, their size and type, and their bonding pattern with the matrix materials [1, 2]. Aluminium, the thirteenth element of the periodic table, has emerged as one of the highly favoured choices from researchers worldwide for selecting the matrix of MMC. Aluminium and its alloys find a special place in automobile and aircraft industries as it helps in reducing its overall weight subsequently cutting down fuel consumption and resulting in lesser carbon footprints. Imparting aluminium alloys with reinforcement particles aids in further enhancing its strength [3]. Generally, hard ceramic materials like Al_2O_3, TiO_2, ZrO_2, SiC, B_4C, ZrB_2,

Recommended for publication by Commission XV - Design, Analysis, and Fabrication of Welded Structures.

✉ Vibhu Singh
vibhusingh101@gmail.com

M. S. Niranjan
mahendraiitr2002@gmail.com

Qasim Murtaza
qasimmurtaza@dce.ac.in

[1] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

Published online: 08 July 2024

TiB_2, TiN, and BN are injected into an aluminium alloy to reinforce its various mechanical properties [4]. From all the above ceramics available, titanium diboride (TiB_2) stands out as one of the predominant choices as it imparts advanced phase strengthening of aluminium alloys since it portrays exceptional characteristics like a high melting point of 2900 °C, less density of 4.52 g/cm^3, enhanced corrosion resistance behaviour, and minimum wear rate [5]. Also, from earlier research work [6], it is widely acknowledged that the addition of rare earth elements (REE) like Er, Sc, Sr, Yb, La, Zr, and Ce or their oxides like CeO_2 or La_2O_3 can significantly modify or enhance the mechanical properties of various aluminium alloys through grain refinement, improving dispersion of other ceramic particles, control of dynamic recrystallisation etc. Furthermore, the addition of rare earth oxide (REO) also helps in expanding various alloying possibilities. However, the utilization of REE or REO has been limited because of their soaring market cost [6, 7]. Different methods known for fabricating MMC include stir casting, squeeze casting, friction stir casting, centrifugal casting, and a range of in situ methods [8]. Nowadays, various state-of-the-art techniques, such as wire arc additive manufacturing, are also gaining attention for fabricating composites [9]. But among these, conventional stir casting stands out as one of the most widely employed techniques because of its simplicity, cost-effectiveness, and versatility.

The limited compatibility of MMCs with conventional welding methods has been a significant barrier to their widespread application, since research on fusion welding techniques for MMCs, including laser welding, electron beam (EB) welding, metal inert gas (MIG) welding, and tungsten inert gas (TIG) welding, remains limited [10]. The following literature highlights some of the research. Garcia et al. [11] performed a comparative study using direct and indirect electric arc MIG welding on Al-1010/TiC/50p composite. Results showed that indirect arc showed better weld uniformity, less TiC dissolution, lessened heat-affected zone, and consistent mechanical strength (182–186 MPa), which concluded that indirect arc is preferable for welding these composites, even with high reinforcement content. Kun et al. [12] conducted an EB welding of TiB_2/ZL101 composite, resulting in defect-free welding seams with a homogeneous dispersion of TiB_2 particles. Welding has improved the mechanical properties of the base alloy. Tensile tests reveal that fractures occur in the base metal, and the elastic modulus of the welded specimens is relatively greater compared to the pre-welded base alloy. Also, under proper parameters, EB welding successfully prevented porosity and deepened the narrow cross-section of the weld. A study conducted by Ureña et al. [13] involved TIG arc-welding of thick sheets of aluminium alloy (AA) 2014/SiC composite, with varying SiC volume fractions of 6%, 13%, and 20%. The research aimed to explore the impact of interfacial reactions between the matrix and reinforcement on the fracture performance of the composite during welding. Tensile results consistently indicated that failure predominantly occurred within the weld metal, exhibiting its lower tensile strength compared to base alloy due to the formation of Al_4C_3, which weakened the interface between reinforcement and matrix and limited the propagation of cracks within the composite particles. Yue et al. [14] fabricated a composite sheet of AA2124/25(vol)% SiC (3 μm) with 3.5 mm thickness and utilized pulsed Nd-YAG laser welding to join the sheets. Results showed that the presence of brittle carbides in the centre of the fusion zone, which is mainly impacted by laser intensity and pulse duration, and by optimizing these parameters, excessive carbide formation, which can be detrimental to mechanical properties, could be mitigated. Welding MMCs using conventional fusion welding techniques like MIG or TIG presents several challenges. These include the segregation of strengthening agents, chemical interactions between reinforcement and matrix, and the appearance of various interfacial phases. Also, the occurrence of porosity in the liquified state or at higher solidification temperature and liquation cracks can develop between the reinforcements and the matrix. These problems are worsened by the high heat input required in the melt pool, leading to the degradation of the MMC's properties [10, 15]. Therefore, solid-state processes are often preferred for the welding of MMCs, among which friction stir welding (FSW) and friction welding (FWs) are mainly significant. Devanathan et al. [16] utilized FSW to join B_4C-reinforced AA5083 MMC, employing tools coated with titanium nitride, aluminium chromium nitride, and diamond-like carbon with varying friction coefficients and the flower pollination algorithm (FPA) was employed for optimization. Results showed that the coating significantly influenced tensile strength, microstructure, and tool condition. Ali et al. [17] performed a study to examine the impact of process parameters on the properties of friction stir welded MMCs comprising AA6061 with various weight percentages of SiC and B_4C particles. Outcomes revealed that the highest tensile strength, reaching 174 MPa, was achieved with composites containing 10% SiC and 3% B_4C. Moreover, the elongation percentage rose with increasing B_4C content but declined with decreasing SiC content. As a development of FSW, the friction stir vibration process, denoted as FSVP, uses controlled vibrations to improve the weld quality on alloys and composites. Abbasi et al. [18] fabricated a surface composite using an Al5052 alloy sheet and SiC nanoparticles via friction stir processing (FSP). A comparative study of the microstructure and mechanical properties between FSP and FSVP, where the workpiece is vibrated perpendicular to the processing line, was conducted. The results indicated that FSVP refined the grain in the stir zone and improved the homogeneity of particle distribution. Also, FSVP specimens

demonstrated better strength, hardness, and ductility than FSP specimens, which increases on increasing the vibration frequency. Barati et al. [19] used FSP and FSVP to AA5052 to fabricate a surface composite with SiC nanoparticles and compared their effects. The results revealed that FSVP significantly improves the microstructure and mechanical properties compared to FSP as it reduces the stir zone's grain size and attains a more homogeneous distribution of SiO_2 particles, resulting in greater strength, ductility, and wear resistance. The literature above highlights FSW as among the most preferred methods for welding MMCs, yet they are disadvantaged by their exposure to defects like cracking, splattering, extreme generation of heat during rotation, creating a turbulent flow of base metal, a lengthy welding procedure, less flexible to fusion welding, fast and extreme wear of the tool, contamination of the joint by remains from the tool, the sizeable downward force required for clamping weld plates, non-preferrable for mass production, inferior efficiency in comparison to fusion welding, demands a high early investment cost, higher energy consumption, rigid process control conditions, or probable material damage from excessive vibrations [20, 21]. Hence, traditionally available welding methods may not always meet the demands of joining MMCs, requiring exploring alternative approaches such as cold metal transfer (CMT) welding. With its precise control over heat input and metal transfer, the CMT technique may present a promising solution for welding MMCs. CMT, a modified and colder version of conventional MIG, can potentially knock off the above-mentioned issues by providing the advantages of 1/3rd heat input, spatter-free material transfer, penetration control to get a proper dilution ratio in bead, and an option to control arc length during welding in comparison to other fusion processes [22, 23]. CMT is an emergent technology that involves dip transfer technology and uses a three-phase welding cycle as follows: (1) In peak current, this phase occurs within a constant voltage environment with high current. The high current is necessary to establish an arc between the filler wire and the base metal, subsequently heating the filler wire for melting. (2) In background current, this phase involves a reduction in the current value to inhibit the transfer of filler metal droplets in the form of globules, promoting a short-circuit transfer between the molten pool and the tip of the filler wire and (3) in short circuit, in this phase, the short-circuit current is further reduced to a lower current level while simultaneously sending a signal to retract the wire from the wire feeder. This action prompts a backward drawing force within the wire feeder, causing the droplet to detach from the filler wire and transfer into the molten weld pool [23]. The above three-step cycle then repeats. By carefully optimizing different process parameters influencing the CMT technique, MMCs can be efficiently joined while minimizing defects such as porosity, solidification cracking, and distortion. Isitan et al. [24] casted two distinct composites using the liquid metallurgy vortex-route process with AA6013 as matrix and nano-Al_2O_3 as reinforcement (0.5% wt. and 1% wt.) and joined via butt welding at two current (110A and 120A) and using welding methods, i.e. CMT and pulse multi-control (PMC). Results showed that nanocomposites with nano-sized reinforcement reduce grain size and improve mechanical properties with a composite having nano-Al_2O_3 reinforcement (0.5%wt) welded at 120 A using the CMT process, achieving the highest tensile strength value. Further, the CMT process allows for better control over the formation of intermetallic compounds at the interface between the matrix and reinforcement, which can significantly influence the different properties of the welded joints. Overall, utilizing the CMT technique for welding MMCs can have excellent potential for different industries, like aerospace, automotive, defence, sports, and marine applications, where welding of lightweight, high-strength materials is in demand [25]. From the above literature, it is evident that though CMT welding has been extensively used in traditional aluminium alloys [26], its use in the welding of MMCs composed of these alloys remains relatively unexplored, representing a significant research gap. Furthermore, fusion welding, specifically CMT welding, of hybrid composites comprising a novel combination of hard ceramic and rare earth oxide reinforcements into the AA-6061 T6 matrix still needs to be explored.

In the present work, a novel hybrid composite has been fabricated by incorporating micro-sized TiB_2 particles and nano-sized REO La_2O_3 particles as reinforcements within an AA6061-T6 matrix using conventional stir casting. The CMT technique is then employed to weld two similar plates of fabricated composites. Response surface methodology (RSM) using central composite design (CCD) is utilized to model and optimize the tensile strength of CMT weld joints in composite plates by considering three critical process parameters: current, travel speed, and gas flow rate. Subsequently, the accuracy of the developed model is validated through experimental verification. Lastly, an in-depth investigation of the microstructural evolution and mechanical properties of the CMT weld joints of fabricated composite has been performed to provide an overall better understanding of the welding process and its effects.

2 Experimental methods

2.1 Materials

In this research, the matrix material employed for the preparation of the composite is AA6061-T6 (Fig. 1a), with its elemental composition shown in Table 1. Hard ceramic TiB_2 powder (Fig. 1b) with 14 µm particle size and 2 wt.% and REO La_2O_3 powder (Fig. 1c) with a particle size ranging

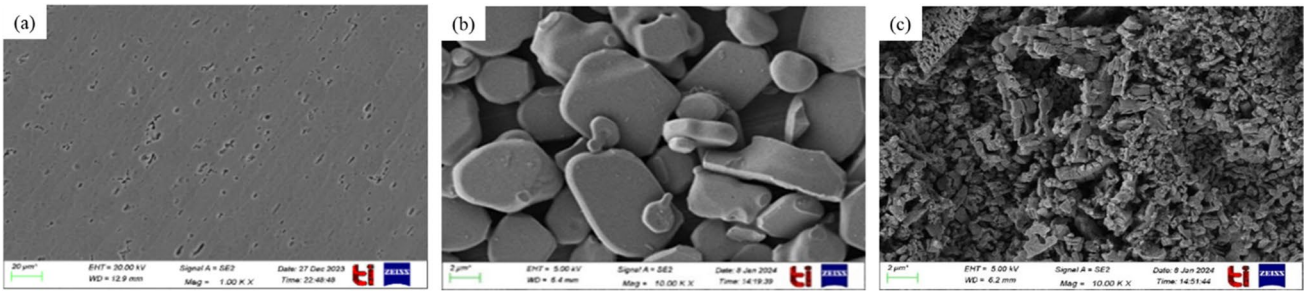


Fig. 1 FESEM image of **a** AA6061-T6, **b** TiB_2 powder (size 14 µm), and **c** La_2O_3 powder (size 40–50 nm)

from 40 to 50 nm and a 0.5 wt.% are selected as reinforcements. Figure 2 displays the XRD peaks of each powder and matrix involved in the fabrication of the hybrid composite.

2.2 Composite plate fabrication

The preparation of the composite follows the methodology outlined in established literature [8, 27]. A solid cylindrical bar of AA6061-T6 weighing 1 kg is utilized as the matrix for preparing the composite. Micro-sized TiB_2 ceramic powder with 2 wt.% and nano-sized La_2O_3 powder with 0.5 wt.% are selected as reinforcements. Figure 3a illustrates the complete methodology involved in fabricating the composite plates. Initially, the AA6061-T6 cylindrical bar is heated within the furnace of stir casting setup, achieving temperatures ranging from 700 to 800 °C for approximately 1 h, transforming it into a molten state. Simultaneously, both reinforcements are preheated at 300 °C in a muffle furnace to remove any possibility of impurities, moisture, or dust. Once the alloy gets melted, a stirrer is inserted into the melt, up to a depth of 1/3rd of the molten surface, to create a vortex. Following the formation of a vortex, both reinforcements are introduced via a feeder positioned at its centre, with TiB_2 powder being added first, followed by La_2O_3 powder. The stirrer is continued at 600 rpm for the next 10 min to assist in the mixing of reinforcement particles and arrest any possibility of agglomeration of particles within the molten matrix. As soon as mixing is achieved, the hot molten mixture is poured into a cuboidal mould (Fig. 3b) and kept untouched for cooling and solidification to achieve a casted composite, as shown in Fig. 3c. Wire EDM process is performed on each fabricated cuboidal composite to get the plates of the required dimension displayed in Fig. 3d.

2.3 CMT welding

Composite plates with dimensions of 110 mm × 60 mm × 3.2 mm have been used for the welding purpose. CMT welding has been carried out using ER5356 filler wire (1.2 mm Dia) with the elemental composition mentioned in Table 1, keeping the torch angle at 90° , CTWD to 13 mm and using pure Argon as a shielding gas. As shown in Fig. 4, pure argon was used as the shielding gas during CMT welding. TransPuls Synergic Advanced Cold Metal Transfer Welding (400TPSi) machine setup shown in Fig. 4 and supplied by Fronius, Austria, has been used for welding. Plates are cleaned with acetone to prepare them for welding and are joined in a square-butt configuration through a single pass, as shown in Fig. 5a.

Process parameters that need to be investigated depend upon the response parameter. After an in-depth review of the literature, it has been determined that the major factors influencing the response, i.e. tensile strength of the weld joint, are current, travel speed, and gas flow rate. To initiate the welding process, a few welding trials were conducted with a range of values for the above-mentioned factors. These values were chosen based on a few previous literatures [28–31] and the constraints associated with parameter selection in the Fronius CMT welding machine setup due to the choice of ER5356 filler wire. Considering these factors, the ranges on which experiments were done are as follows: current (A): 110, 115, 120, 125, 130, 135, 140, 145, and 150; travel speed (mm/s): 6, 8, 10, 12, 14, and 16; and gas flow rate (L/min): 12, 14, 16, 18, 20, and 22. After welding with various parameter combinations, each weld bead was inspected based on specific visible

Table 1 Elemental composition of matrix and filler wire

Materials	Si	Cu	Fe	Zn	Mg	Ni	Mn	Cr	Ti	Sn	Pb	Al
AA6061-T6	0.79	0.19	-	0.07	0.98	0.03	0.17	0.045	0.03	0.025	0.024	Rem
ER5356	0.03	0.01	0.15	0.01	4.83	-	0.14	0.11	0.09	-	-	Rem

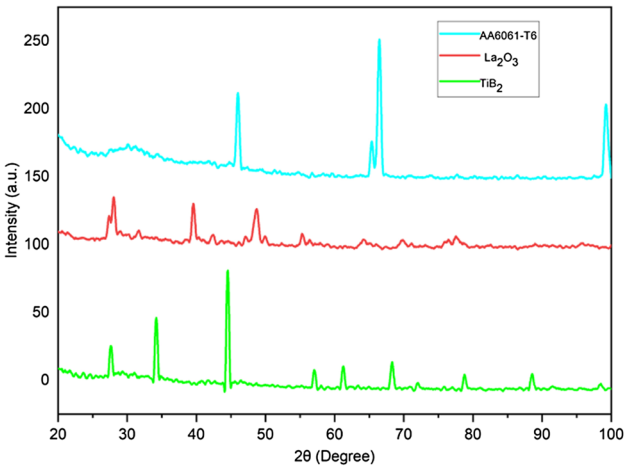


Fig. 2 XRD peaks of reinforcement powders and matrix

criteria, including spatter amount, penetration depth, base plate distortion, arc stability, and surface porosity, as shown in Fig. 5b. Following inspection, the optimal levels for each control factor were finalized, as detailed in Table 2.

2.4 Development of design matrix for CMT welding

RSM is a multivariable statistical tool used to improve outcomes by developing, analyzing, and optimizing the model. It establishes a relationship between multiple input process variables and desired results. It is one of the most popular and fittest optimization tools as it helps achieve the experimental goal with a minimum number of runs compared to other tools. This method significantly improves processes by examining the effects of different input factors and their interactions on the response variable. Its main advantage

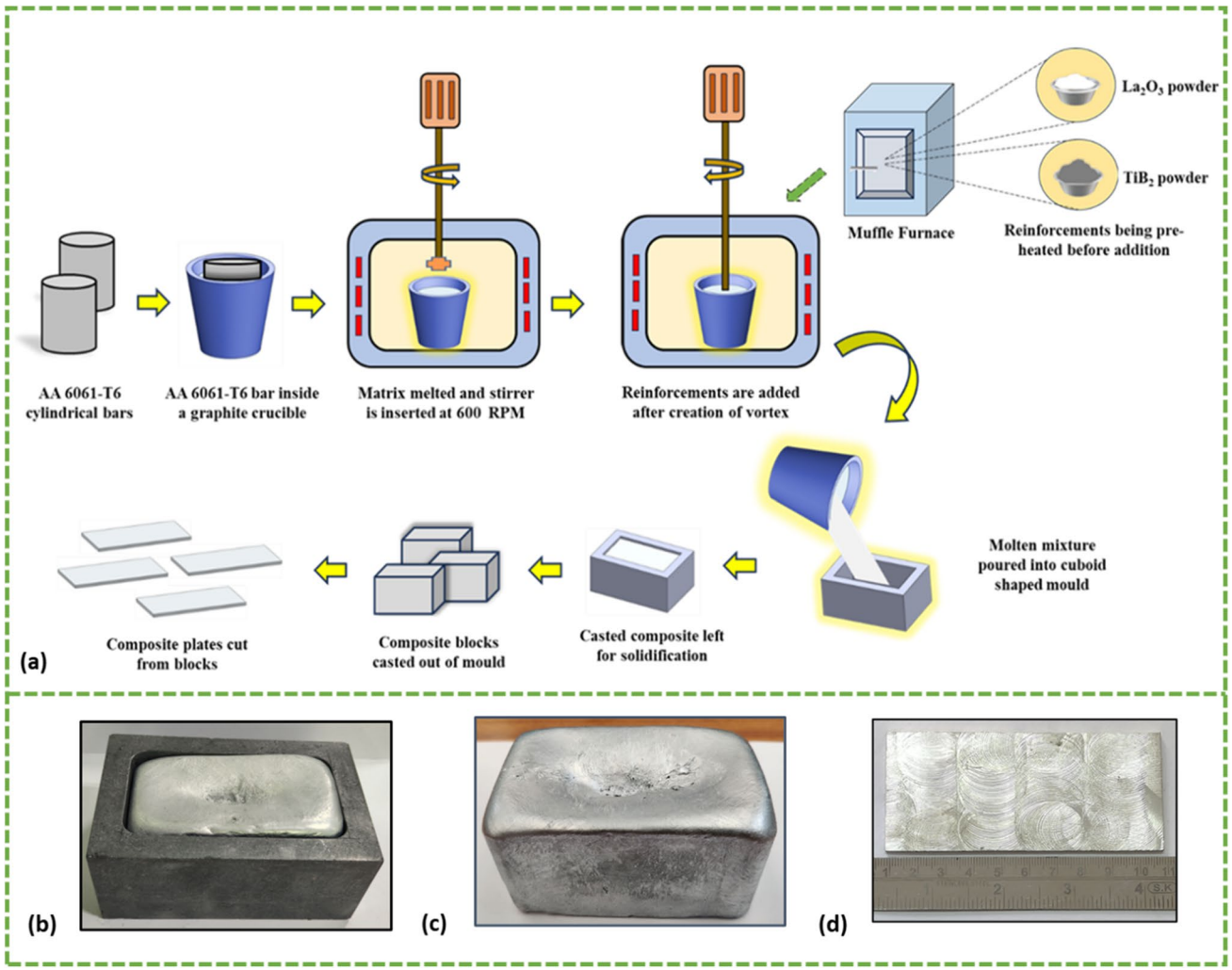


Fig. 3 **a** Steps involved in the fabrication of composite plates. **b** Casted composite solidifying in the graphite mould. **c** Fabricated composite. **d** Composite plate cut from the casting

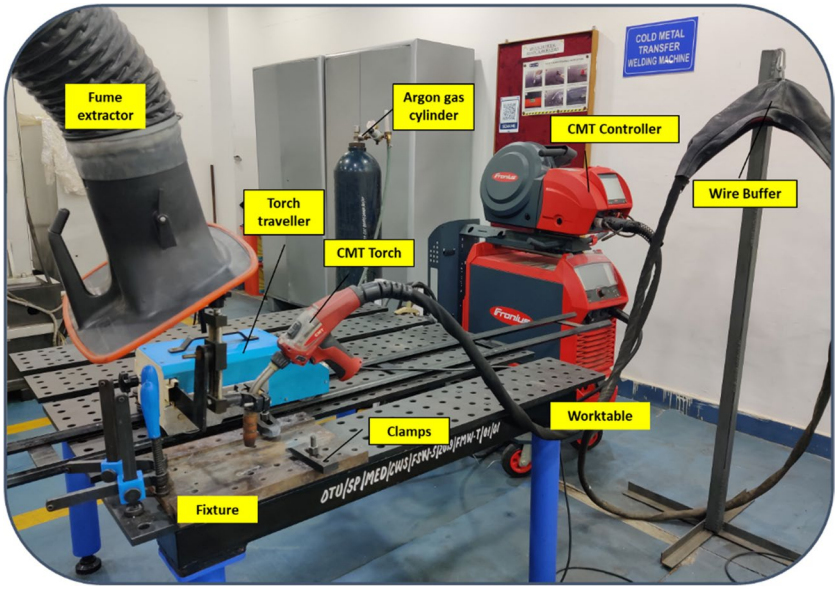


Fig. 4 Lab setup of CMT machine

is identifying optimal process parameters while minimizing experimental runs. By minimizing the number of runs, RSM reduces the resources, time, and costs related to the experimental phase of process optimization, making it specifically helpful in various engineering and industrial fields like mechanical, chemical, medical, or manufacturing, where experimental costs can be high [32, 33]. Here, response surfaces are graphically represented to interpret the interactions between control factors and their resulting outputs. Generally, two types of RSM design are available, central composite designs (CCD) and Box-Behnken designs (BBD), among which CCD is one the popular and favoured designs developed by Box and Wilson in 1951 since it is suitable to fit a quadratic model and is of great use in the successive investigation as it provides to the development of previous factorial trails by additional axial and centre points [34, 35]. In this work, the rotatable design of CCD (shown in Fig. 6) has been selected as it contributes to consistently predictable variance at every point which is at an equal distance from the centre of the design matrix, thereby providing a quality prediction. The flowchart employed for optimization using the RSM is displayed in Fig. 7. Design Expert software has been used to design the CCD matrix having three control factors with five levels, each with $-\alpha, -1, 0, +1$, and $+\alpha$. A total of 20 runs is conducted in CCD, which is calculated using the formula: Total runs $= 2^p + 2*p + N$ centre points (p = input variables, N = no of repetition of centre value) using $\alpha = 2$.

Quadratic response surface models are constructed to analyze tensile strength based on experimental data and ANOVA. The resulting response function is represented by Eq. (1), where Y = tensile strength.

$$Y = \varphi(current, travel\ speed, gas\ flow\ rate) \tag{1}$$

The second-order response, incorporating the K factor-surface model equation, is defined in Eq. (2) as follows [35]:

$$Y = \alpha_0 + \sum_{i=1}^{K} \alpha_i X_i + \sum_{i=1}^{K} \alpha_{ii} X_i^2 + \sum_{i<j} \sum_{j=2}^{K} \alpha_{ij} X_i X_j + \epsilon \tag{2}$$

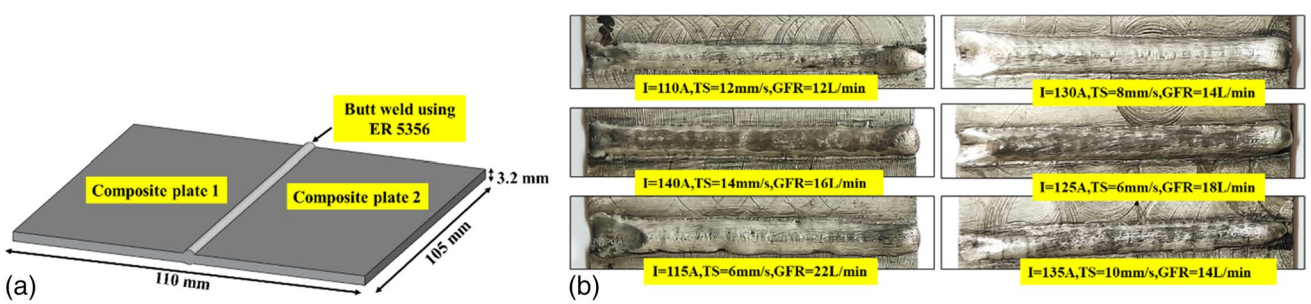


Fig. 5 **a** Schematic diagram of square butt weld joint of composite plates. **b** Trial weld beads of composite plates

Table 2 Process parameters and their levels selected for CMT welding of composite plates

Process parameters	Unit	Level				
		−2	−1	0	+1	+2
Current	A	125	130	135	140	145
Travel speed	mm/s	8	10	12	14	16
Gas flow rate (GFR)	L/min	12	14	16	18	20

Here, Y denotes the output response, X_i represents the linear terms, X_i^2 defines the quadratic terms, X_iX_j signifies the interaction terms, a_0, a_i, a_{ii}, and a_{ij} are constants associated with the input parameters, and ϵ represents the fitting error. Figure 8a shows a total of 20 runs for CMT welding with square butt joint configurations, which is performed as per the process parameters outlined in Table 3. Once all composite plates are welded, each plate is subjected to a wire EDM process to get the required specimen for different tests, as illustrated in Fig. 8b.

2.5 Porosity and dilution

Porosity always stands out as a prevailing defect during welding, resulting in a decreased strength of the welded joint by weakening atomic bonds within the material. This defect is marked by the existence of gases or voids within the melted weld pool. In the present study, the calculation of porosity percentage at two distinct locations across the fusion zone has been conducted using a digital camera and an inverted metallurgical microscope. The microscope employs a specific formula (Eq. 3) to analyze the percentage of porosity within the weld bead [28]

$$Porosity\ percentage = \left\{ \frac{(Total\ feature\ area\ X\ 100)}{Total\ area} \right\} \quad (3)$$

In fusion welding, dilution denotes the ratio of filler wire present in the weld pool relative to the parent material. This ratio is the proportion of the penetration area (Ap) to the total area (Ap + Ar) resulting from a single bead deposit, generally expressed as a percentage [36].

2.6 Microstructural characterization

After CMT welding, the plate is cut using the wire-EDM technique to get different sizes of specimens for various microstructural and mechanical tests. The specimen has been cold moulded using the epoxy-hardener combination for the optical microscope. After mounting, the specimen is subjected to various dry and wet polishing to get a mirror-like polish on the surface. Finally, it is subjected to Keller's reagent for etching. Weld microstructure has been observed using an OLYMPUS optical microscope and ZEISS Gemini 1 Sigma-500 field emission scanning electron microscopy (FESEM) setup equipped with an energy-dispersive spectrometer (EDS). Electron backscatter diffraction (EBSD) analysis is performed on the weld specimen after it is electropolished using methanol (70%) and nitric acid (30%) at a voltage of 13 V for a time of 13 s to get an insight into the grain size. X-ray powder diffractometer (XRD) BRUKER D8 Advance machine has been employed to identify different chemical compounds formed during the welding. Post tensile test, the fracture surface is analyzed using a JEOL JSM-6610 LV scanning electron microscope (SEM) provided with EDS.

2.7 Mechanical test

The microhardness test is achieved using a Duramin-40 M1 machine, set to a hardness scale of 0.5 HV with a dwell time of 10 s between consecutive indentations. Specimen for the test has been prepared as per ASTM E384 standards with indentions taken along a linear line of the weld bead at a space of 1 mm each. The Tinius

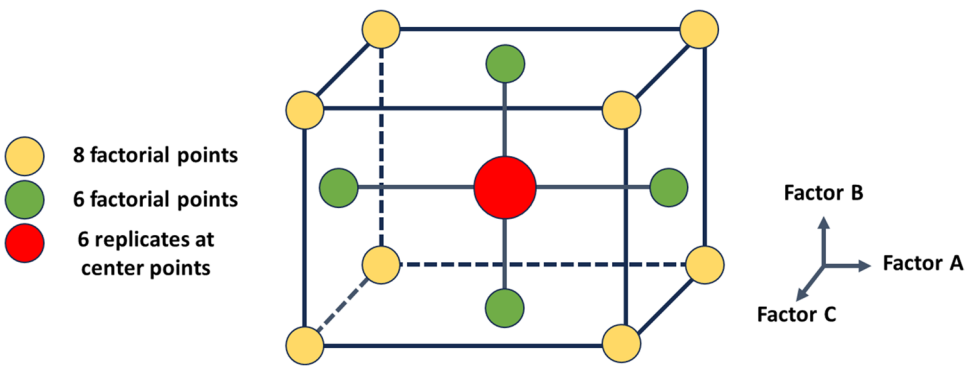


Fig. 6 Position of points in the CCD matrix

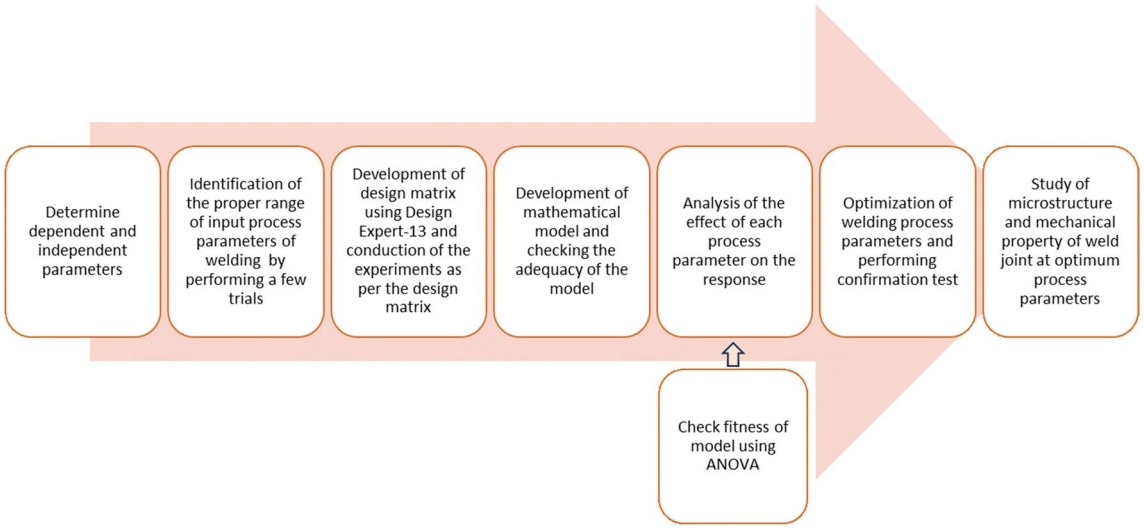


Fig. 7 Flowchart for optimization using RSM

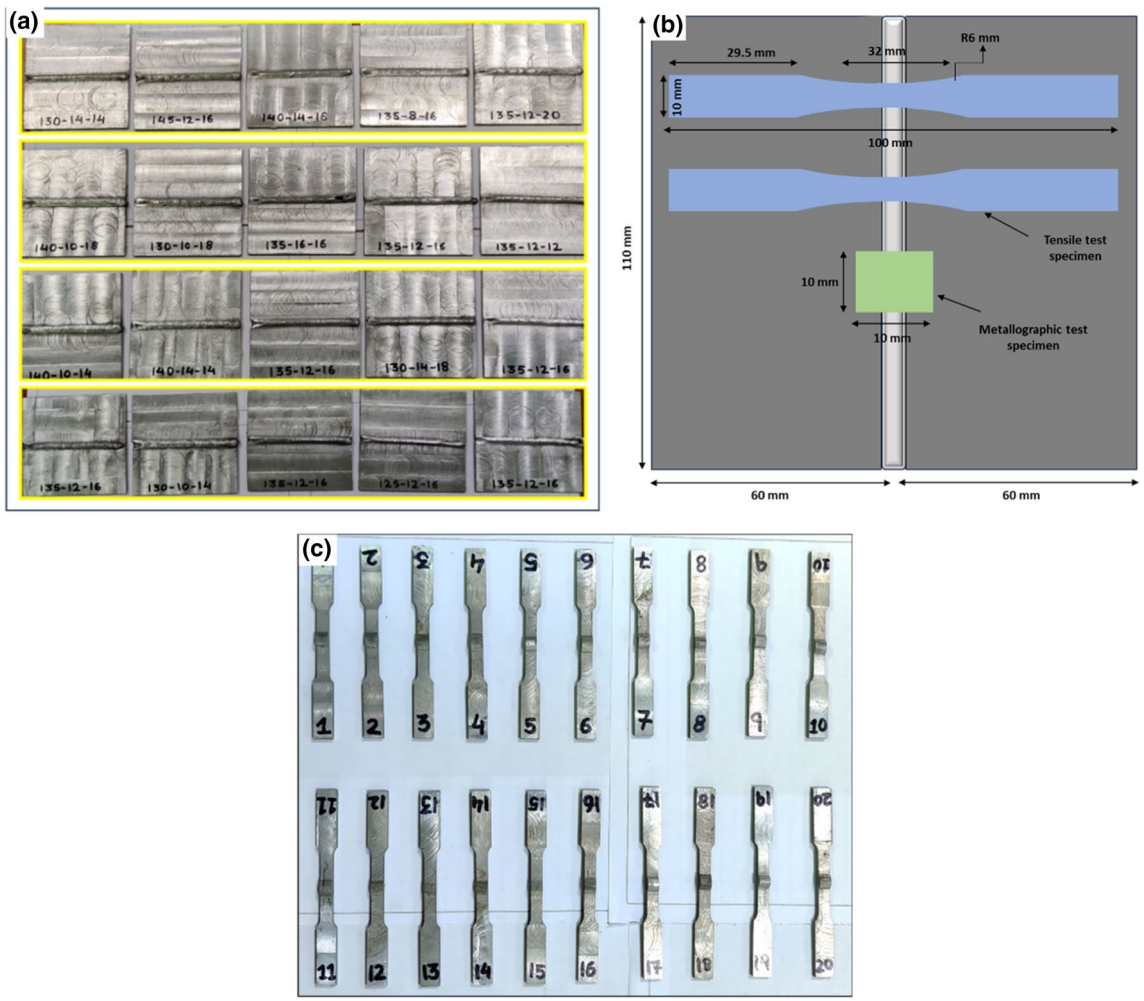


Fig. 8 **a** CMT-welded composite plates. **b** Schematic representation of test specimen cutoff from weld joint. **c** Tensile specimens of the CMT weld joint of composites

Table 3 Input factors for CMT welding and corresponding response

Std	Run	Current (A)	Travel speed (mm/s)	GFR (L/min)	Tensile strength (MPa)
3	1	130	14	14	143.28
10	2	145	12	16	161.24
8	3	140	14	18	152.17
11	4	135	8	16	155.8
14	5	135	12	20	149.89
6	6	140	10	18	163.03
5	7	130	10	18	144.29
12	8	135	16	16	143.61
15	9	135	12	16	158.07
13	10	135	12	12	153.09
2	11	140	10	14	166.04
4	12	140	14	14	154.58
16	13	135	12	16	158.9
7	14	130	14	18	141.9
20	15	135	12	16	159.62
18	16	135	12	16	159.02
1	17	130	10	14	146.08
17	18	135	12	16	160.51
9	19	125	12	16	129.42
19	20	135	12	16	158.03

Olsen H50KS universal testing machine has been used to conduct tensile tests on two samples prepared from each welded specimen according to ASTM E8 standards, as shown in Fig. 8c. The average value of these tests is then used for further evaluation.

3 Results and discussion

3.1 Statistical analysis

Twenty welded specimens are prepared by joining composite plates in square butt joints, and tensile tests are performed on each of them. Table 3 displays the tensile strength of all weld specimens. ANOVA assesses process parameters' significance and model adequacy. The lists of the sum of squares (SS), degrees of freedom (DOF), *F*-value, and *P*-value are mentioned in Table 4. *F*-statistic is only a ratio of two variances-variation between factors to variation within the factors with higher values denoting greater significance of that particular input parameter. The *P*-value symbolizes and demonstrates the probability at a confidence level of 95%, such that a *P*-value < 0.05 denotes the significance of the model term. In contrast, a *P*-value exceeding 0.1 indicates insignificance [37, 38]. Model adequacy is assessed based on R^2, adjusted R^2, and predicted R^2 values. Table 4 shows that the model displayed an *F*-value of 283.94 and a *P*-value < 0.005, indicating its significance. Consequently, the lack-of-fit *F*-value of 0.63 suggested that most residual variations originated from pure errors rather than a lack of fit. The significant regression and non-significant lack of fit affirmed the model fitted with the experimental data. Notably, terms such as A, B, C, A^2, B^2, C^2, and AB are significant, whereas AC and BC are insignificant. Additionally, the R^2 value of 0.9961 and adjusted R^2 value of 0.9926, both nearing 1.0, indicated a strong fit with the experimental data. The predicted R^2 value of 0.9872, differing by less than 0.2 from the adjusted R^2 value, implied the model's suitableness for

Table 4 ANOVA of full quadratic model for tensile strength of CMT weld joint of composite

Source	Sum of squares	df	Mean square	*F*-value	*P*-value	
Model	1560.18	7	222.88	389.39	< 0.0001	Significant
A-Current	959.61	1	959.61	1676.50	< 0.0001	
B-Travel speed	168.29	1	168.29	294.01	< 0.0001	
C-GFR	14.04	1	14.04	24.54	0.0003	
AB	36.68	1	36.68	64.08	< 0.0001	
A^2	294.20	1	294.20	513.99	< 0.0001	
B^2	136.14	1	136.14	237.84	< 0.0001	
C^2	88.93	1	88.93	155.37	< 0.0001	
Residual	6.87	12	0.5724			
Lack of fit	2.39	7	0.3417	0.3816	0.8793	Not significant
Pure error	4.48	5	0.8954			
Cor total	1567.05	19				
Std. dev	0.7566		**R^2**	0.9956		
Mean	152.93		**Adjusted R^2**	0.9931		
C.V. %	0.4947		**Predicted R^2**	0.9883		
			Adeq precision	74.4189		

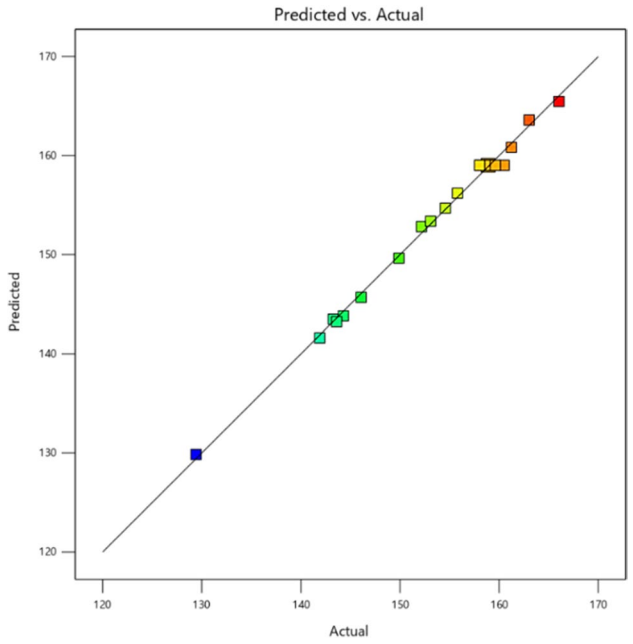


Fig. 9 Predicted vs. actual graph for tensile strength of CMT joint

further predictions. The adequate precision, exceeding 65.1709 (> 4), indicated an adequate signal, affirming the tensile strength model's significance and precision.

Once the adequacy and reliability of the models are confirmed, insignificant terms are eliminated without compromising the accuracy of the models. Through regression analysis, the quadratic equation for predicting tensile strength is derived and is represented by Eq. 4:

$$\begin{aligned} \textit{Tensile Strength} &= 159.03 + 7.74A - 3.24B - 0.9369C \\ &\quad - 2.14A * B - 3.42A^2 - 2.33B^2 - 1.88C^2 \end{aligned} \tag{4}$$

Here, A, B, and C represent current, travel speed, and gas flow rate, respectively. A positive sign preceding the terms denotes that the term works to create a better response, whereas a negative sign indicates that the term works against and reduces the overall outcome [39]. The quality of the developed model is estimated using the correlation coefficient value. Figure 9 illustrates the predicted vs actual graph in which actual and predicted responses clustered closely around the $y = x$ line, affirming the significance of the regression equation model. Consequently, the ANOVA outcome is considered reliable. Figure 10a and b illustrate 2D and 3D graphical representations of the regression equation, showing the relationship between tensile strength, current, and travel speed while maintaining a constant shielding GFR of 16 L/min.

3.2 Effect of process parameters on response

3.2.1 Single effect of current

The welding current influences the weld bead's metallurgical and mechanical properties as changes in current can result in transformations of the structure of the weld zone at the micro level, affecting its mechanical properties. It is

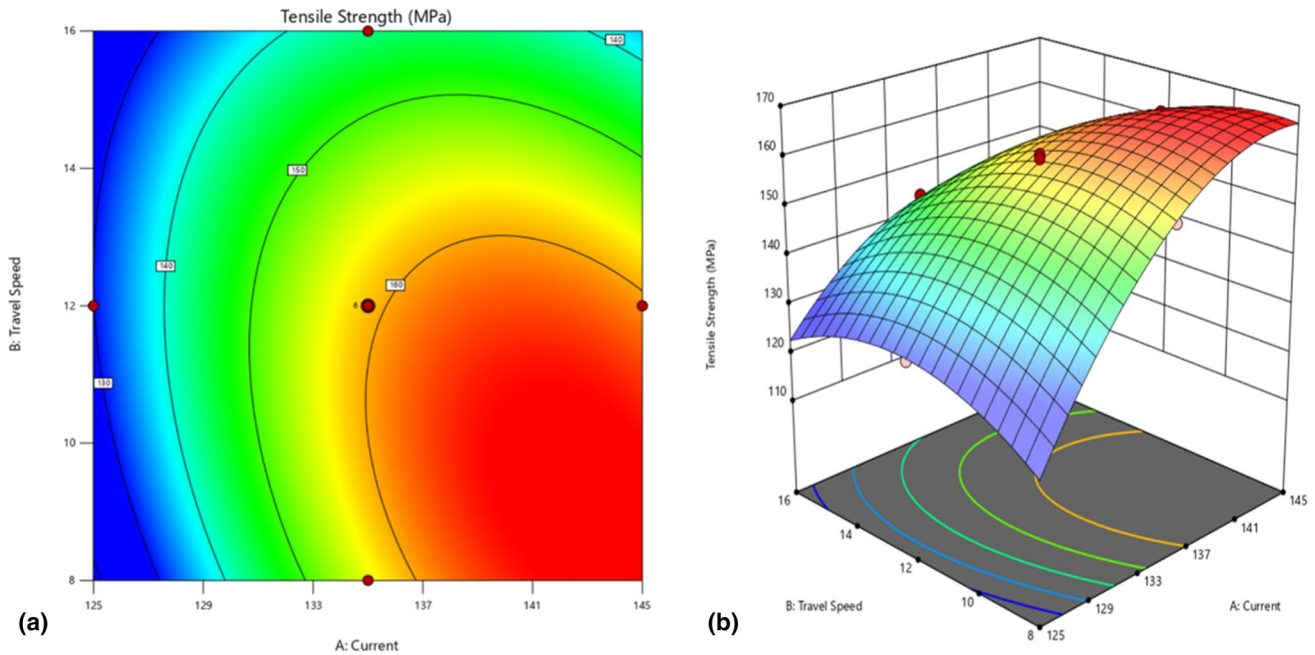


Fig. 10 Interaction between current and travel speed. **a** 2D contour plot, **b** 3D response surface plot

evident that change in the current tends to modify the tensile strength of the weld sample by directly influencing heat input into the weld pool. When the current is lower, there is lessened heat input, as confirmed by Eq. 5 [40].

$$Q = k\frac{VI}{v} J/mm \tag{5}$$

where Q is the heat input (J/mm), k is the thermal efficiency, V is the arc voltage (V), I is the current (A), and v is the travel speed (mm/s). Therefore, due to the decreased heat input, the filler wire melts less. This results in inadequate penetration, particularly with ER5356 filler wire, which has lower silicon (Si) content. This lower Si content reduces the wire's fluidity, which might contribute to the formation of pores when the filler wire is introduced into the weld. Thus, the welded joint exhibits lower tensile strength [41]. As the heat input increases, the filler wire melts more, as indicated by Eq. 5, and therefore, the weld bead's geometry, i.e. width and penetration, generally increase. This increase in bead geometry can also be justified by Eq. 6 [42], which relates arc force (F) and current:

$$F = \frac{K}{2} I^2 \tag{6}$$

where K is a coefficient, and I defines the current. According to Eq. 6, as the current increases, the heat input and the value of F also increase. Typically, an increase in F facilitates greater arc penetration, enhancing weld penetration. Therefore, increasing the welding current leads to the opposite effect, resulting in an upward trend in tensile strength. The temperature also rises with increasing heat input due to higher current in CMT welding. This increase in temperature influences the absorption of hydrogen. Generally, as temperature increases, the solubility of hydrogen in the aluminium matrix also increases, following Sievert's Law [43], according to which the total volume of hydrogen in the liquid metal ([H]) is proportional to the partial pressure of hydrogen in the arc (P_{H2}), represented as $[H] = K\sqrt{P_{H2}}$. At high temperatures, the favourable distribution of hydrogen atoms into the aluminium lattice happens due to increased kinetic energy, facilitating the increased dissolving of hydrogen atoms into the aluminium matrix. This increased porosity can reduce the tensile strength of the bead at a higher current [44].

3.2.2 Single effect of travel speed

The speed at which the welding torch travels is crucial in determining the tensile strength of weld specimens, as it directly impacts the weld bead geometry and, consequently, the mechanical properties of the weld joint. Decreasing the speed may result in excessive heat being applied per unit length to the base material, as indicated by Eq. 5, which shows an inverse relationship between travel speed and heat input. This excessive heat also causes the excessive melt of filler material and possibly results in the collapse of the weld bead and the formation of defects through more penetration. On the other hand, increasing the speed can lead to incomplete fusion, as less heat will be shared with the base material per unit length, resulting in reduced penetration and weakening of the weld joints [45].

3.2.3 Single effect of GFR

A suitable amount of shielding gas is significant to keep the weld pool stable, prevent excessive splattering, and regulate the heat during welding. Results have shown that when the GFR is too low, it fails to adequately protect the weld area, resulting in defects like porosity that can weaken the weld's tensile strength. As GFR increases and reaches an optimum value, it leads to a reduction in porosity, ensuring consistent deposition of filler material, uniform heat distribution, and less spatter, therefore providing better tensile strength comparatively. However, further increasing the GFR beyond the optimum level can actually be detrimental in shielding a weld pool and contribute to porosity, as excessive gas flow can create turbulence around the weld pool. This disrupts the smooth flow of shielding gas and allows atmospheric air to mix in, which is readily absorbed by the molten weld metal. Upon solidification, these gases are released and form porosity, leading to less tensile strength [46, 47].

3.2.4 Interaction effect of current and travel speed

Current and travel speed play an interactive role in influencing the tensile strength of the CMT weld joint of composite plates. Increasing current generally leads to increased heat input into the weld zone, as per Eq. 5. More heat allows for deeper penetration and better fusion between the base material and filler metal, potentially increasing strength. However, excessive current has negative consequences, such as too much heat can cause excessive melting of the filler wire and base material, leading to the collapse of the weld bead. Travel speed is inversely related to heat input as per Eq. 5. A slower travel speed allows for more heat input as the torch dwells on the joint for a longer duration, whereas a faster travel speed does the opposite. Materials might require higher currents and slower travel speeds for proper penetration. It is observed that the tensile strength maximizes around higher current values and lower travel speed [40–42, 45].

3.3 Optimal parameters

Design Expert Software provides numerical optimization options to determine an optimal combination of variables

Table 5 Maximum and minimum limit of process parameters

Factors	Units	Goal	Min limit	Max limit
Current	A	In range	125	145
Travel speed	mm/s	In range	8	16
GFR	L/min	In range	12	20
Tensile strength	MPa	Maximize	129.42	166.04

that satisfy defined requirements. The primary objective of this optimization is to achieve the maximum response while ensuring that all variable properties are satisfied simultaneously. The software employs five goal options to create desirability indices: maximize, minimize, target, in range, and equal to. Table 5 displays the criteria of optimal parameters used in the experiment. This experiment used numerical optimization to optimize the response variable, i.e. tensile strength. The single optimization results indicated that the maximum tensile strength of CMT welded joints could be achieved when the current is set at 141.637 A, the travel speed is set at 8.6 mm/s, and the gas flow rate is 13.75 L/min. The maximum value of tensile strength achieved at these values comes out to be 166.247 MPa, as displayed in Fig. 11. However, due to the constraints of the CMT machine setup, these values are rounded to the nearest whole numbers, i.e. a current of 142 A, a travel speed of 9 mm/s, and a gas flow rate of 14 L/min.

3.4 Confirmation test

Confirmation is performed using a single combination of optimal parameters to validate the empirical equation related to current, travel speed, and GFR. If the average observation falls within this prediction interval, it confirms the model's accuracy. Three confirmatory runs were carried out during the welding process to validate the model, and the resulting averages were recorded, as shown in Table 6. The validity test results show that confirmation results closely match the predicted values, revealing that the designed model is accurate.

3.5 Microstructure characterization of welded specimen at optimum process parameters

3.5.1 Porosity and dilution

Figure 12a represents a good-quality bead, showing no sign of solidification cracking and minimal presence of pores. The challenge of solidification cracking has commonly been prominent in fusion welding of MMC, but CMT welding at its optimal setting has effectively addressed this issue, resulting in fewer defects. The porosity in the fusion zone of the weld bead is calculated at two locations, as displayed in Fig. 12b, using the formula mentioned in Eq. 3. The total porosity percentage for both locations is determined to be 1.82% and 1.76%, resulting in an average value of approximately 1.79%. During welding, the affinity of hydrogen gas towards liquid aluminium matrix increases, which might result in pores formation within the fusion zone of the CMT welding [44]. The dilution of the welded joint of composite plates comes out to be 61.26% using the dimensions shown in Fig. 13a and b. This shows a complete penetration of the filler wire and is a balance between excessive and insufficient dilution since excessive dilution can lead to issues

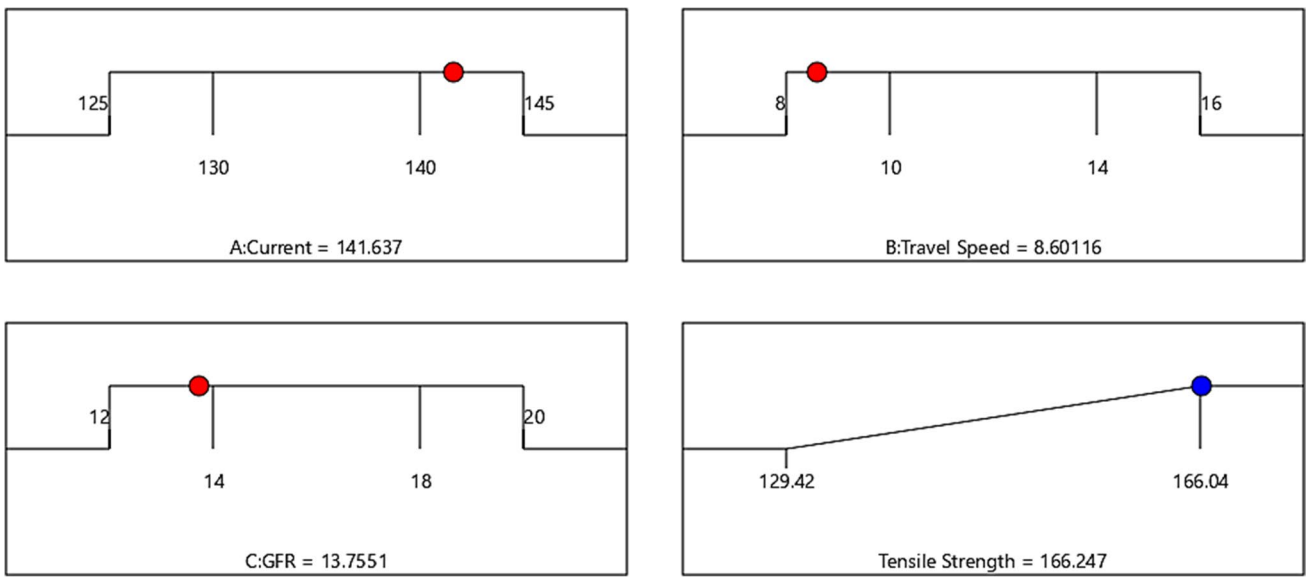


Fig. 11 Ramp function graph showing the optimal value of each process parameters

Table 6 Predicted and experimental values at optimum parameters

Run no	Current (A)	Travel speed (mm/s)	GFR (L/min)	Tensile strength (MPa)		
				Predicted	Experimental	Error %
1	142	9	14	166.247	165.04	0.726
2	142	9	14	166.247	165.73	0.310
3	142	9	14	166.247	166.81	0.338

such as reduced mechanical strength, changes in chemical composition, and the formation of undesirable microstructures, while insufficient dilution may result in inadequate fusion between the base plate, leading to defects such as lack of penetration or incomplete joint formation [28, 41].

3.5.2 Optical and FESEM analysis

Figures 14 and 15 show the optical and FESEM images of three zones formed during CMT welding of composite plates, i.e. fusion zone (FZ), heat-affected zone (HAZ), and base material (BM), which is a fabricated composite in this study. FZ shows finer and equiaxed dendrites formed at a high current value of 142A. At a high current value, a faster solidification rate occurs, thereby promoting the formation of smaller grains [48]. However, these dendrites are distributed non-uniformly throughout the zone. At HAZ, coarse columnar crystals are visible. In any welding system, the solidification structure mainly depends on the ratio of the thermal gradient of the liquid–solid line (G) and the solidification growth rate (R). The value of this ratio increases from the centre of the FZ toward the solid–liquid boundary. Now, at a higher current value, the heat input in the weld zone increases, leading to a higher thermal gradient near the fusion line. However, the growth rate will be minimal, resulting in a lower cooling rate; hence, grain growth will be slow, and coarser grain will be formed in HAZ [48, 49]. BM shows the non-uniform distribution of both reinforcements within the volume of AA6061-T6, which substantially impacts the composites' properties, such as mechanical and fracture behaviour. The grayscale contrast discovered in the FESEM images corresponds to the α-Al phase, while the darker areas reveal the presence of secondary phases such as Mg_2Si, $Al_{12}Mg_{17}$, and reinforcement TiB_2. Also, a few accumulations of micro TiB_2, along with a few precipitates, are visible along the boundary. During solidification, the microstructures of composites include a solid solution of aluminium and an inter-dendritic structure of eutectic aluminium–silicon [50]. The reinforcement particles are constrained toward refined α-Al grains as the composites solidify. Due to this, more refinement of α-Al grains occurs, and the TiB_2 and La_2O_3 particles serve as nuclei on which the α-Al grains solidify, providing resistance against the increasing α-Al phase during the solidification process of casting [50]. Figure 16 displays the EDS analysis at the fusion line, in which different peaks conform to the presence of various elements. The presence of titanium (Ti) and lanthanum (La) suggests the incorporation of reinforcement materials within the matrix of the fabricated composite, accompanied by lesser peaks signifying the presence of magnesium (Mg) and silicon (Si).

3.5.3 Electron backscatter diffraction (EBSD)

The inverse pole fig (IPF) map shown in Fig. 17a illustrates different zones in the CMT-welded composite, specifically focusing on the region near the fusion line of the weld bead.

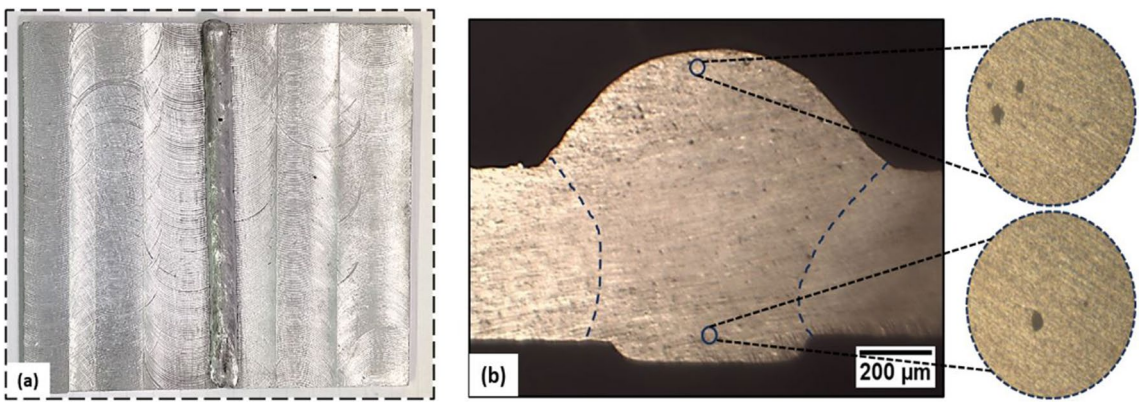


Fig. 12 **a** Top view of CMT weld bead at optimum welding parameters. **b** Location for porosity measurement

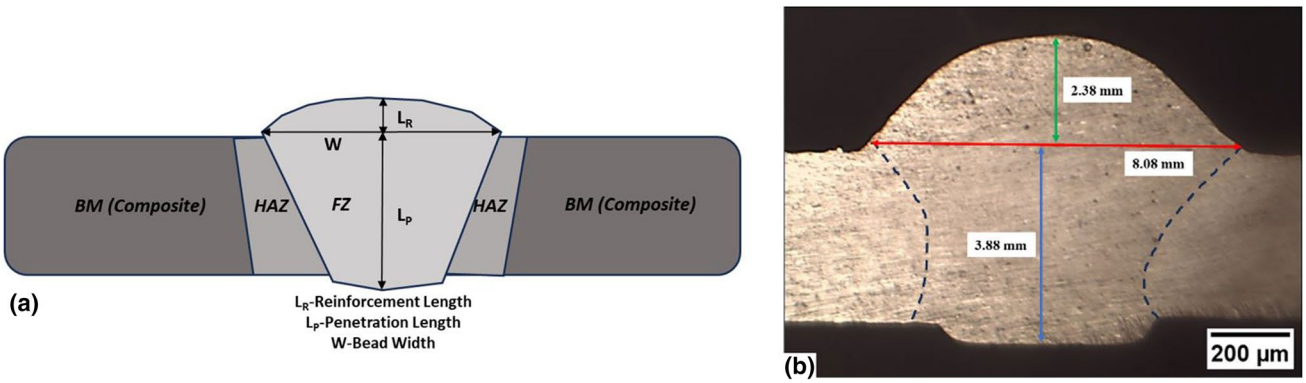


Fig. 13 **a** Geometry of CMT weld bead, **b** Dilution of weld bead

On the left side of the fusion line, a cluster of finer grains ranging in size from approximately 3 to 9 µm is observed. In contrast, the right side of the fusion line exhibits comparatively larger grains with a size range of approximately 12 to 76 µm. Moving towards the base material further increases the grain size, ranging from ~ 95 to 237 µm in comparison to the fusion zone. Figure 17b displays the boundary map where sub-grain boundaries, low-angle grain boundaries (LAGB), and high-angle grain boundaries (HAGB) are represented with red, green, and blue lines, respectively. The observation reveals a higher prevalence of sub-grain boundaries (< 5°) in comparison to LAGB (5–15°), while HAGB (> 15°) has the highest density among all. The microstructural variations during the welding process with ER5356 filler wire are apparent in Fig. 18a. Also, as per Fig. 17a, the presence and distribution of reinforcement particles within the aluminium matrix are visible, with some TiB_2 particles located at grain boundaries near the

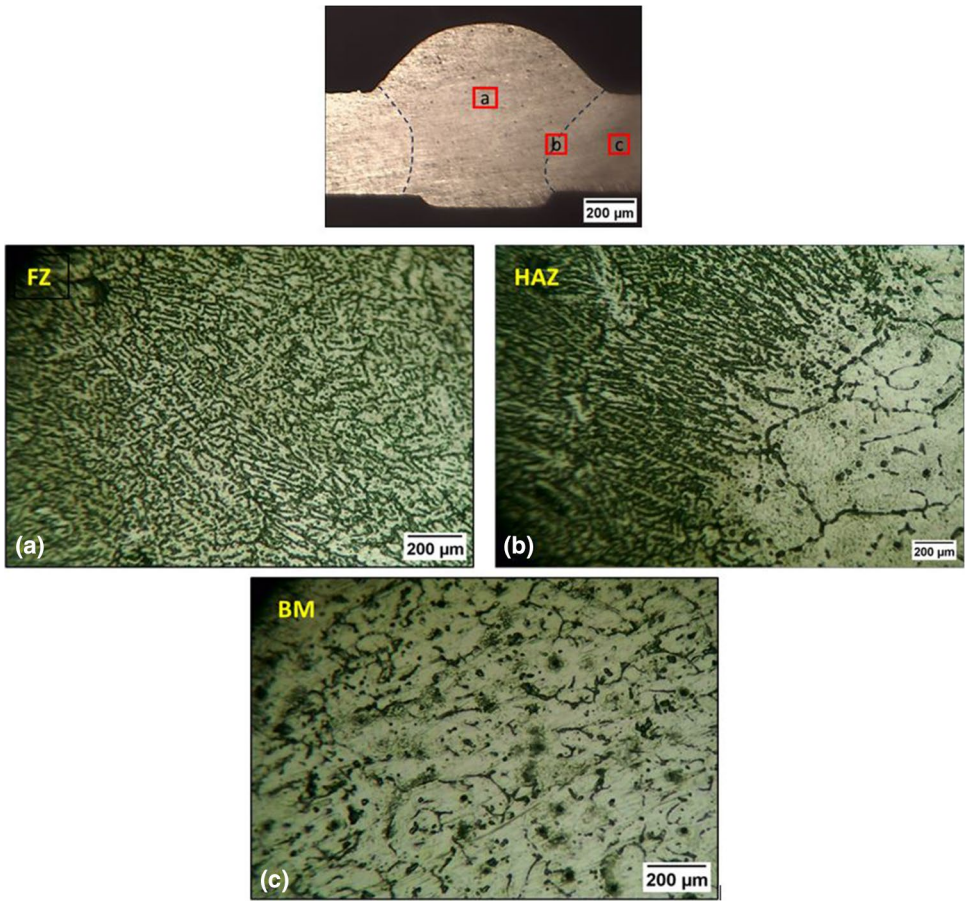


Fig. 14 Optical images of **a** FZ, **b** HAZ, and **c** BM

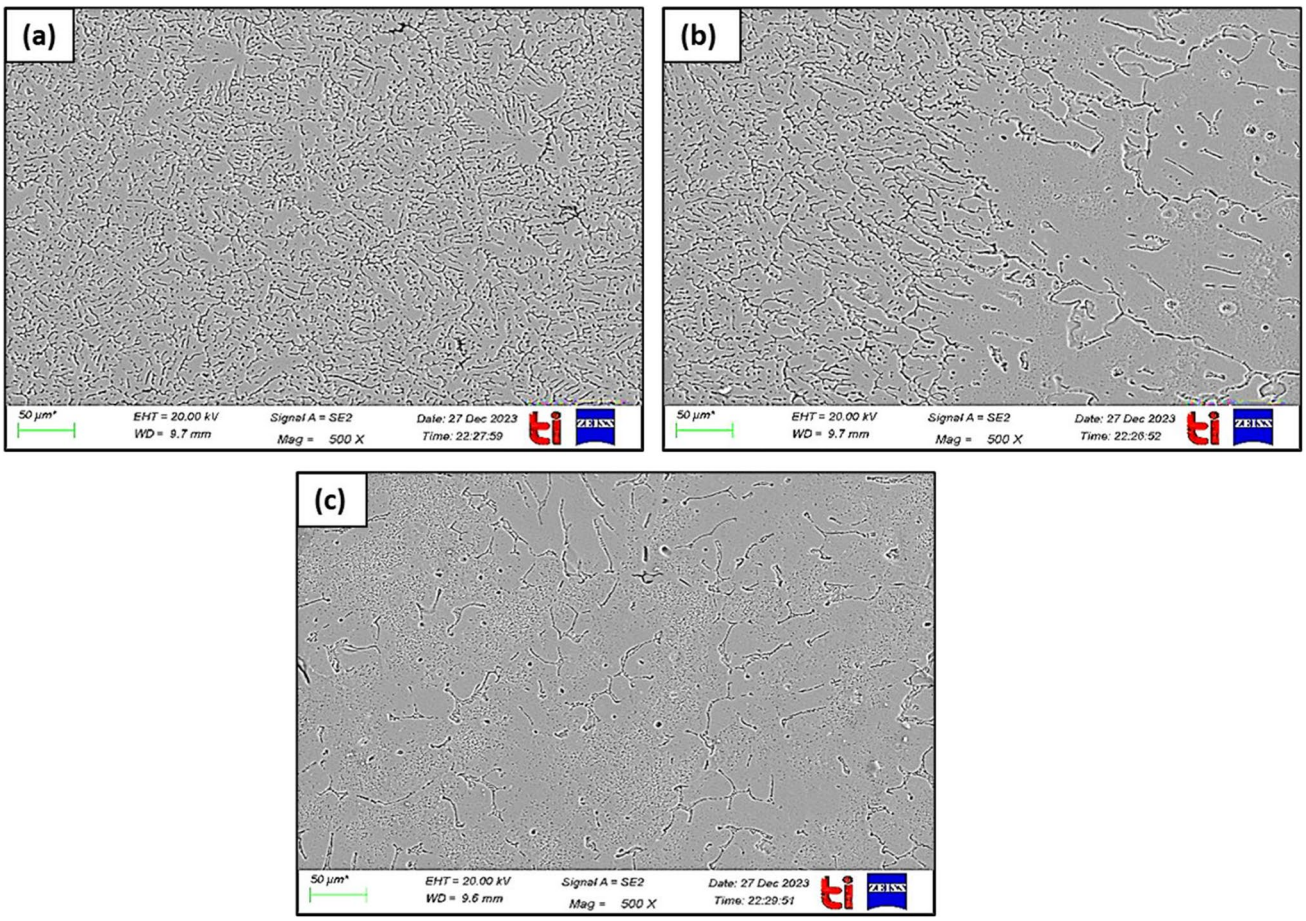
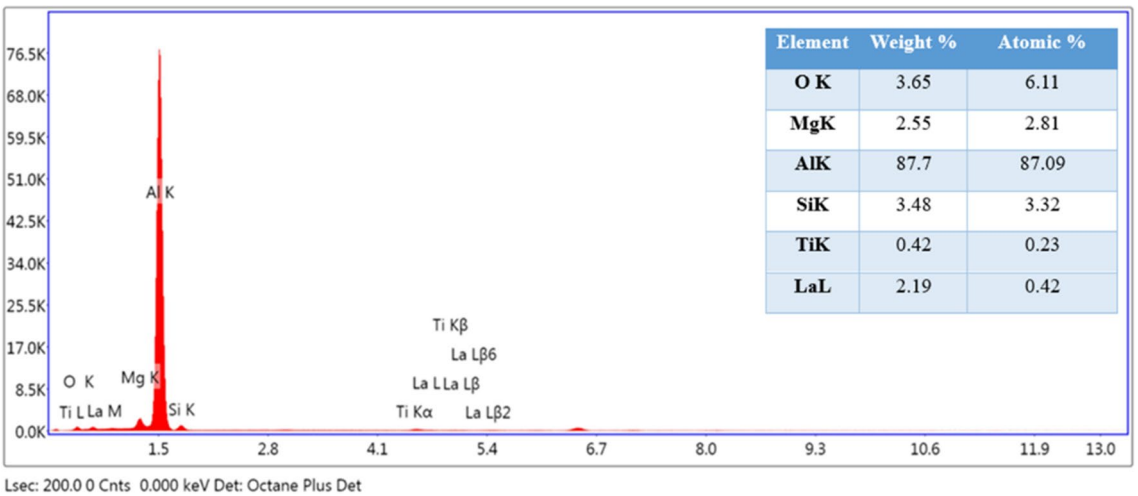


Fig. 15 FESEM images of **a** FZ, **b** HAZ, and **c** BM

HAZ region. This type of distribution is called necklace-type, as shown in Fig. 18b, and is a characteristic of MMCs produced via stir casting, which induces weak interface bonding between reinforcement particles and matrix in particle-segregated zones [51]. This is confirmed by the research of Sun et al. [52], which showed that a distinctive necklace-type distribution is shown when SiC particles are imparted into AZ91D composites.



Element	Weight %	Atomic %
O K	3.65	6.11
MgK	2.55	2.81
AlK	87.7	87.09
SiK	3.48	3.32
TiK	0.42	0.23
LaL	2.19	0.42

Fig. 16 EDS analysis at fusion line of weld bead

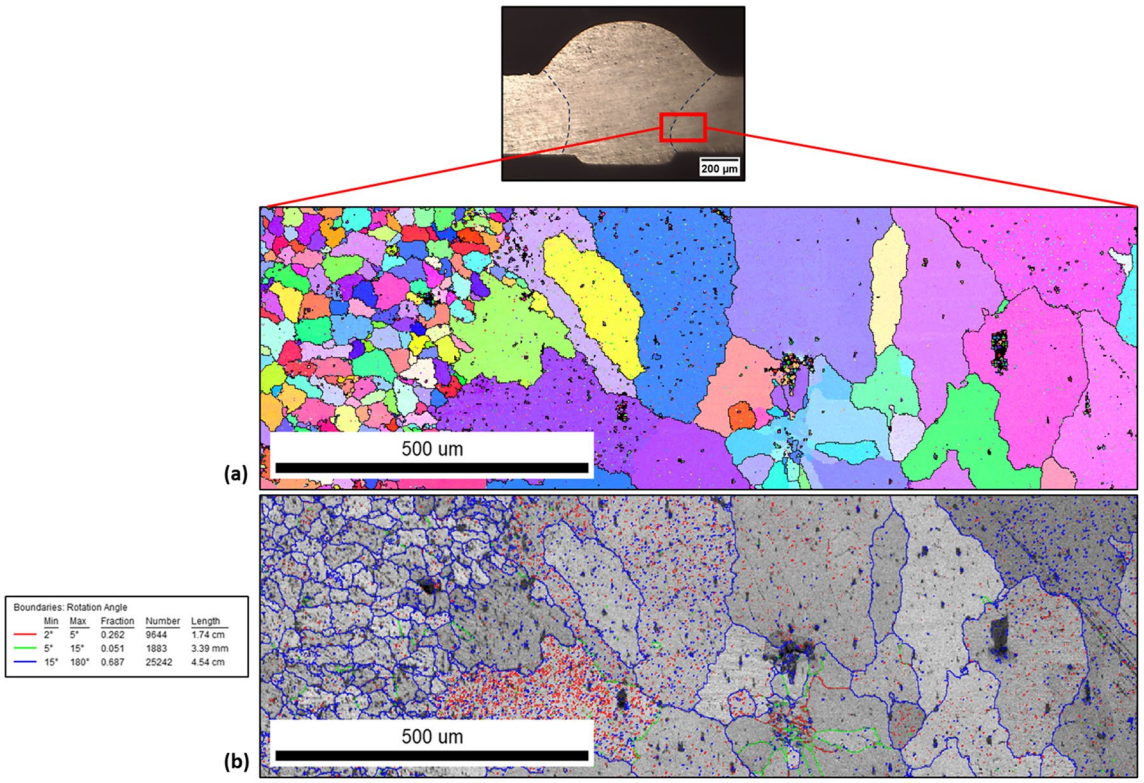


Fig. 17 EBSD analysis of the weld bead near fusion line. **a** Colour inverse pole figure (IPF). **b** Grain boundary map

3.5.4 XRD analysis

In Fig. 19, XRD peaks detected in the fusion zone of the welding reveal the existence of reinforcement materials such as TiB_2 and La_2O_3 within the aluminium matrix, with aluminium exhibiting the most evident peak intensity. Additionally, XRD analysis identified several intermetallic compounds in the weld zone. These compounds form during CMT welding due to the interaction between the filler metal and the dilution of the composite. Major compounds observed include $Al_{3.21}Si_{0.47}$, $Al_{12}Mg_{17}$, Mg_2Si, TiO_2, Al_3Ti, and $Al_{2.12}La_{0.88}$. The rationale behind the formation

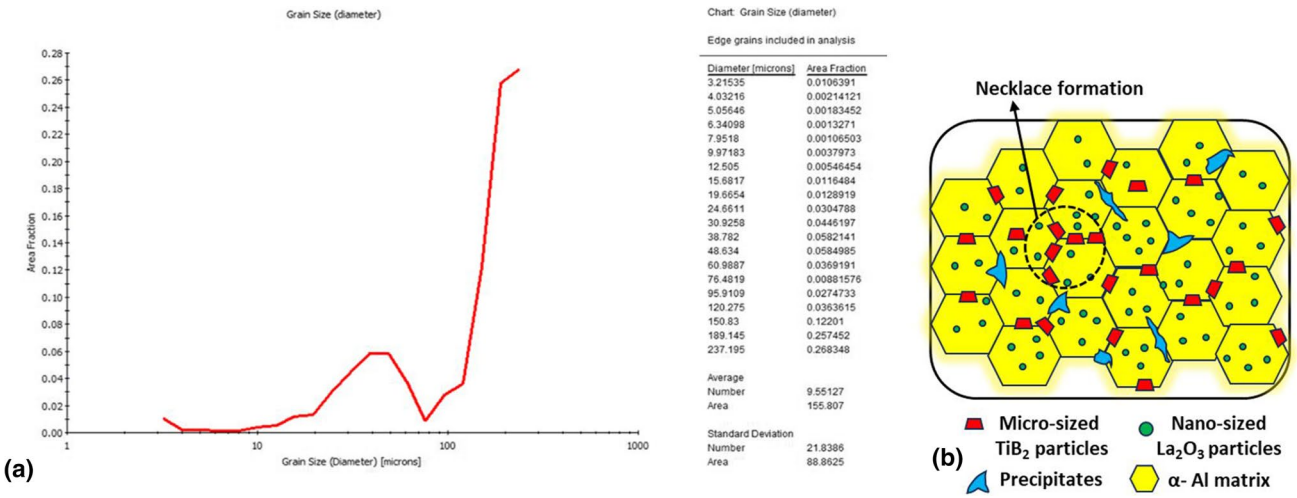


Fig. 18 **a** Distribution of grain size. **b** The schematic diagram illustrates the formation of microstructure in a composite

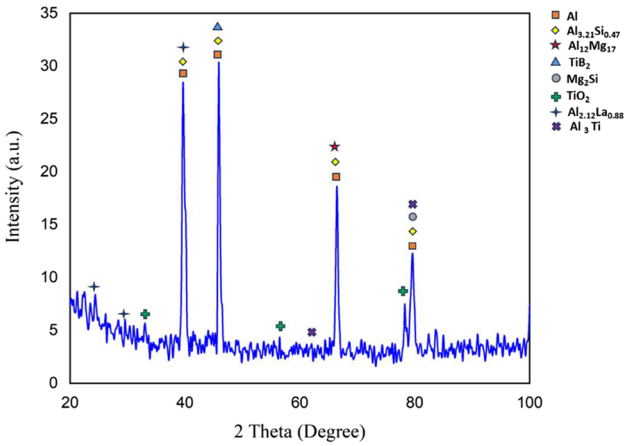


Fig. 19 XRD analysis at the fusion zone

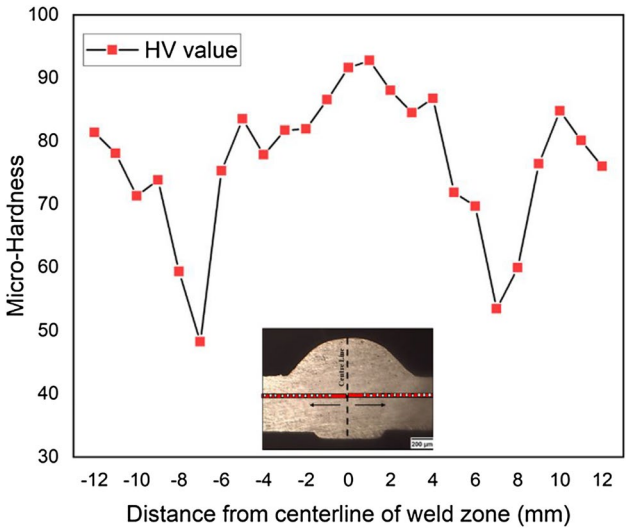


Fig. 20 Average hardness distribution across the weld joint

of intermetallic compounds, $Al_{12}Mg_{17}$, lies in the fact that magnesium (Mg) and aluminium (Al) exhibit complete mutual solubility in the molten state. During the solidification process, a eutectic reaction occurs, leading to precipitation ($Al_{12}Mg_{17}$), which possesses brittle and hard properties. Moreover, during the solidification process of molten Mg-Si within the weld bead region, the presence of silicon (Si) is attributed to diffusion and recrystallization between Si and Mg elements, leading to the formation of Mg_2Si, which is recognized for its favourable mechanical properties, including high hardness, lower density, and a high melting temperature [53]. These findings are consistent with the study conducted by Lakshminarayanan et al. [54], who examined the effect of welding methods (MIG, TIG and FS) on the mechanical properties of AA6061 aluminium alloy and discussed the impact of alloying elements. The appearance of hard and brittle-natured intermetallic compounds such as Al_3Ti and TiO_2 improves the composite's mechanical properties. But at the same time, they pose a challenge in CMT welding due to their tendency to weaken the HAZ [55].

3.6 Microhardness

Figure 20 illustrates the average microhardness results measured at 12 positions across the cross-section of the welded specimen, in which it is observed that the weld joint exhibits the highest microhardness, approximately 92.7HV in FZ. This microhardness is notably greater than the HAZ (48.25HV) and the BM (84.76HV). XRD results indicated the formation and presence of intermetallic compounds such as Mg_2Si and $Al_{12}Mg_{17}$. These intermetallic, along with reinforcement particles, provide additional nucleation sites. The increased number of nucleation sites leads to the formation of more grains and grain boundaries, creating a finer grain structure. These additional grain boundaries inhibit dislocation movement, thereby enhancing the hardness of the weld zone compared to the base composite. EBSD results also confirm this refinement of grains in the FZ. Also, the refinement of grains helps in improving the hardness of the weld zone further comparatively, as confirmed by the Hall–Petch relationship [56] $HV = HV_0 + K_H . d^{1/2}$, where HV_0 and K_H are the constants associated with the hardness measurement and the material, and d is the grain size. HAZ display the lowest hardness, which is mainly attributed to microstructural alterations, such as grain coarsening and recrystallization phenomena, along with the formation of combined precipitates resulting from the thermal fluctuations induced by the welding process [57]. Compared to the HAZ, composites show much higher hardness due to hard ceramic TiB_2 particles [58]. Additionally, the introduction of TiB_2 and La_2O_3 particles enhances the dislocation density around these reinforcement particles during solidification through the Orowan strengthening mechanism [58]. This high dislocation density increases the resistance to dislocation motion, thereby inhibiting plastic deformation. Furthermore, the presence of nano-sized La_2O_3 particles decreases voids within the base alloy [59]. Moreover, an effective bond between the reinforcement and the matrix facilitates the transfer of applied loads from the soft AA6061-T6 matrix to the strong TiB_2 reinforcement, consequently improving the composite's resistance to indentation. Pazhouhanfar et al. [60] validated an increase in the microhardness of the Al6061 matrix as the volume fraction of TiB_2 reinforcement increased. Similarly, Sharma et al. [6] indicated a 33.80% increase in the hardness of the Al6061 with the addition of 2.5 wt.% of REO cerium oxide. Furthermore, the formation of the hard and brittle Al_3Ti and TiO_2 intermetallic compound, as evident from XRD peaks, is another contributing factor to the enhanced hardness of hybrid composites.

3.7 Tensile strength

The tensile test demonstrated that the welded specimen's ultimate tensile strength at optimum parameters is 166.81 MPa, marking a decrease of 25.9% compared to the unwelded hybrid composite's tensile strength of 225.18 MPa. Figure 21a shows that the specimen is broken from the HAZ, indicating that the FZ and BM are stronger than the HAZ under uniaxial tensile loading. The mechanical characteristics of CMT-welded joints primarily depend on defects and metallurgical changes occurring during the welding process. The reason FZ has more strength than HAZ is evident from EBSD, which shows its finer microstructures compared to HAZ, leading to an increase in its strength as per the Hall–Petch effect [60]. It is well-accepted that the increased tensile strength of base materials, i.e. composite, has been brought about by imparting two reinforcements (TiB_2 and La_2O_3) in the matrix, which is responsible for showing various effects within the base matrix to enhance its strength and reducing its ductility like Orowan strengthing, the difference in coefficient of thermal expansion, and the shift of load from matrix to reinforcing agents accompanied by grain refinement. So, in general, the distribution of both the strengthening agents is crucial to increase the strength of the composite [6, 58, 60, 61].

Figure 21b displayed the fracture morphology of the welded specimen after tensile testing, and Fig. 22 demonstrates the mechanism involved in the fracture of CMT-welded joints in composite plates. The fracture region of the specimen is in the HAZ, which is also identified as the weakest area in the hardness test. The fracture displayed a combination of ductile and brittle failures, as both dimples and trans-granular cleavage facets are visible, with brittle failure being the dominant one. The presence of round and elliptical dimples suggests a ductile failure that has elongated in the loading direction, whereas cleavage planes indicate a brittle failure [62]. EDS analysis of the fracture surface, shown in Fig. 23, also confirms the presence and involvement of TiB_2 and La_2O_3 during the fracture of the welded specimen. The fractography of HAZ shows the brittle failure occurred due to the presence of two reinforcements within the composite. However, during welding, HAZ is subjected to a thermal cycle, which leads to modifications in the microstructure, such as grain growth, recrystallization or dissolution, or coarsening of strengthening Mg_2Si precipitates [60]. These changes often lead to coarser and less favourable microstructure compared to the BM and FZ, making the HAZ comparatively weaker. Also, the occurrence of voids or pores during welding initiates cracks that propagate and may result in failure during the tensile test [40].

3.7.1 Post-weld heat treatment (PWHT) of tensile specimen

PWHT is a controlled process involving reheating a welded material to a specific temperature, maintaining it for a set duration, and gradually cooling it down. PWHT is widely recognized for reducing residual stress and enhancing ductility and toughness. When a combination of ductile and brittle failure occurs, with brittle failure predominating, it becomes essential to shift towards a mode of failure where ductility is dominant, as ductile failure is often preferred in structural applications where early warning signs before complete failure are necessary, as it provides better resilience and is less susceptible to abrupt catastrophic failure. In contrast, brittle failure leads to sudden and unexpected fracturing once the material reaches its ultimate tensile strength, which is less desirable [63].

PWHT can enhance ductility, thereby shifting the failure mode from brittle to ductile during tensile testing of welded specimens. In this study, post-welding, composite

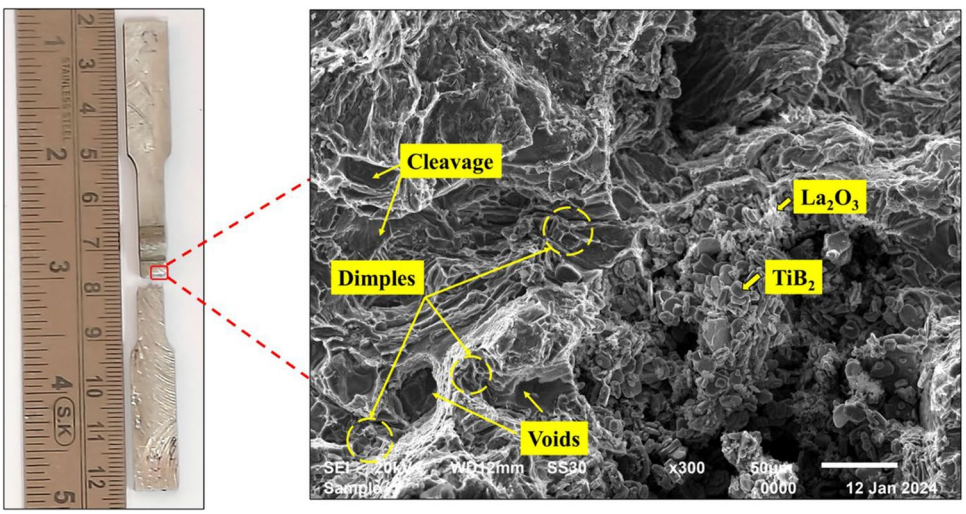


Fig. 21 **a** Tensile specimen of weld joint after the test. **b** SEM analysis of fractured tensile specimen

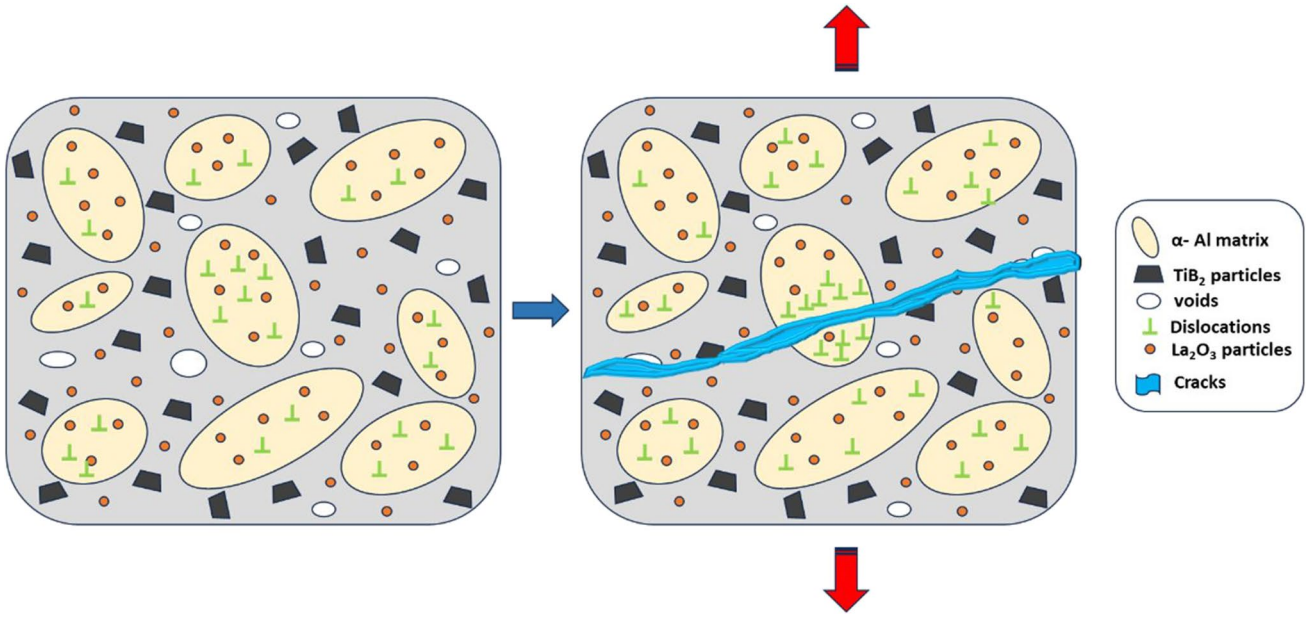
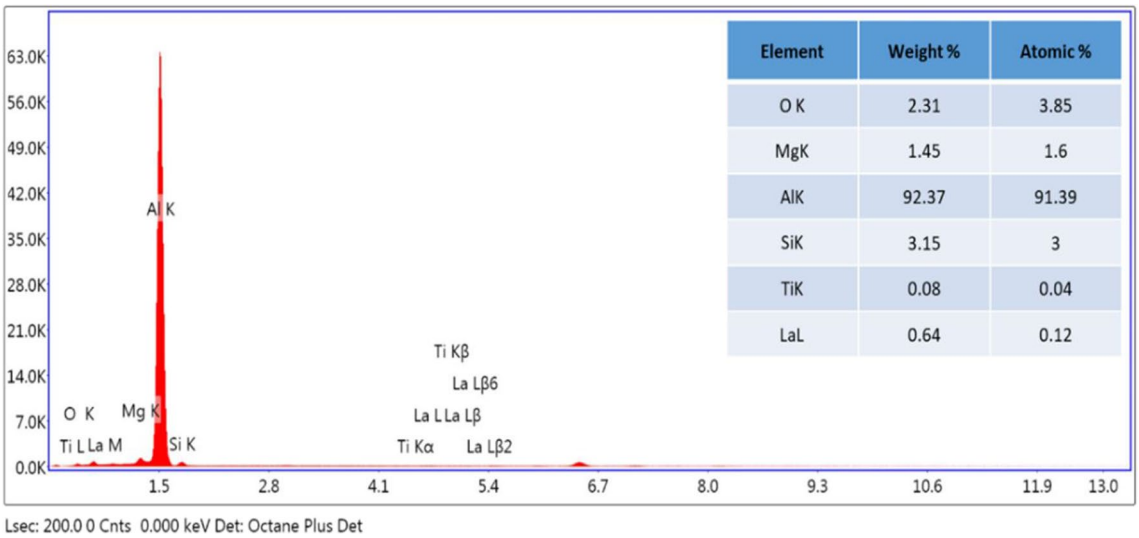


Fig. 22 The basic mechanism involved during the fracture of CMT welded joint of composite plates

plates underwent T6 heat treatment to induce significant microstructural changes. Initially, the specimens underwent solution heat treatment at 520 °C for 2 h in an industrial muffle furnace, followed by quick quenching in water. After that, ageing is conducted at 180 °C for 5 h, followed by cooling in atmospheric air-cooling. Following this, heat-treated welded specimens are subjected to uniaxial tensile testing. Results displayed in Fig. 24a indicate that following T6 heat treatment, the plasticity of the welded composite specimen increased, leading to improved ductility and elongation. Specifically, the elongation increased by 17.80% compared to the non-heat-treated weld specimen, accompanied by a slight increase in strength of 4.13%. Also, fractographic analysis of the heat-treated specimens in Fig. 24b showed that the elongated dimple density increases, with few transgranular



Element	Weight %	Atomic %
O K	2.31	3.85
MgK	1.45	1.6
AlK	92.37	91.39
SiK	3.15	3
TiK	0.08	0.04
LaL	0.64	0.12

Fig. 23 EDS analysis of fractured tensile specimen

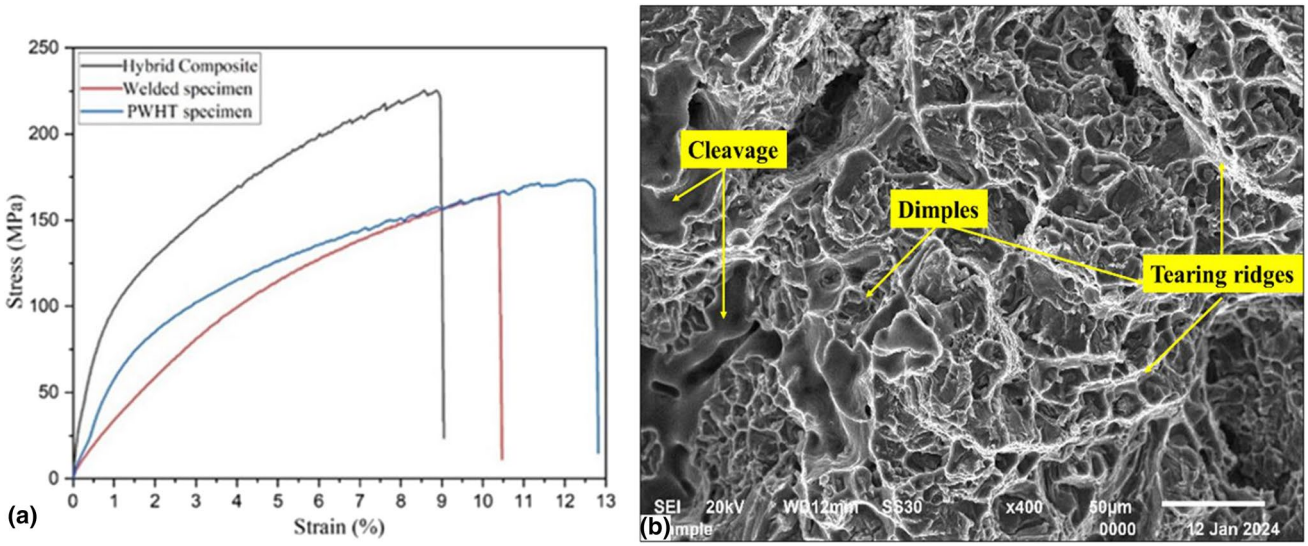


Fig. 24 **a** Stress–strain graph for different tensile specimens. **b** Fractography of tensile specimen post-heat treatment

cleavages, observed due to the presence of hard reinforcement (TiB_2) within the matrix. The T6 heat treatment process may help facilitate the diffusion of alloy elements and enhance the uniformity of microstructures [64, 65]. Overall, the dominant failure mode shifted from brittle to ductile with the application of heat treatment. Li et al. [66] reached a comparable conclusion when applying a T6 heat treatment to an Al/Si/Cu/Mg aluminium alloy containing 1.5 wt.% SiC particles as reinforcement. The composite displayed increased tensile strength and elongation compared to the base alloy.

4 Conclusions

In this study, the process parameters of CMT weld joints of a novel AA6061-T6/2 wt.% TiB_2/0.5 wt.% La_2O_3 hybrid composite fabricated through stir casting are optimized to achieve maximum tensile strength. Quadratic statistical models are formulated to analyze the effect of current, travel speed, and gas flow rate on the tensile strength. Also, the study thoroughly evaluates the influence of these process parameters on weld formation quality. The key conclusions from the experiment are:

- CMT welding of the hybrid composite with an AA6061-T6 matrix and 2 wt.% TiB_2 and 0.5 wt.% La_2O_3 as reinforcements, fabricated via conventional stir casting, was successful, with all experimental runs producing continuous, uniform bead profiles with negligibly detectable visual defects.
- ANOVA indicates that current has the most significant influence on the tensile strength of CMT welds (61.31%), followed by travel speed (10.78%) and gas flow rate (0.89%). Using the desirability function, the optimal input parameters for maximizing tensile strength are current at 142 A, travel speed at 9 mm/s, and gas flow rate at 14 L/min.
- EBSD analysis, along with optical and FESEM imaging, indicates grain refinement in the fusion zone (FZ) compared to the heat-affected zone (HAZ) and base material (BM), where the HAZ exhibits columnar dendritic grains. EDS and XRD analyses confirm the presence of TiB_2 and La_2O_3 within the matrix, along with several intermetallic compounds such as $Al_{3.21}Si_{0.47}$, $Al_{12}Mg_{17}$, Mg_2Si, TiO_2, Al_3Ti, and $Al_{2.12}La_{0.88}$.
- Vickers's microhardness indicates that the FZ exhibits the highest hardness, while the tensile test of weld specimens reveals a 25.9% decrease in strength compared to the unwelded BM (composite), with the HAZ identified as the weakest region.
- The tensile specimens exhibit a primarily brittle behaviour, characterized by cleavage facets and few shallow dimples and tear ridges. Post-weld heat treatment results showed increased ductility indicated by more pronounced dimples and tear ridges and fewer cleavage facets in the fracture analysis.

Given the advantageous characteristics of the CMT welding sources, this technology may emerge as a promising method for welding MMCs, deserving further research and exploration.

Author contribution Vibhu Singh: conceptualization, methodology, investigation, validation, formal analysis, writing—original draft, writing—review and editing; M.S. Niranjan: supervision, writing—review and editing, validation; Qasim Murtaza: supervision, writing—review and editing, validation.

Data availability All data generated or analyzed during this study are included in this article.

Declarations

Competing interests The authors declare no competing interests.

References

1. Zhou MY et al (2020) Progress in research on hybrid metal matrix composites. J Alloy Compd 838:155274. https://doi.org/10.1016/j.jallcom.2020.155274
2. Singh B, Kumar I, Saxena KK, Mohammed AK, Khan MI et al (2023) A future prospects and current scenario of aluminium metal matrix composites characteristics. Alex Eng J 76:1–17. https://doi.org/10.1016/j.aej.2023.06.028
3. Aniruddha MV, Aravindan S, Singh IP (2015) Nano and hybrid aluminium based metal matrix composites: an overview. Manuf Rev 2:15. https://doi.org/10.1051/mfreview/2015018
4. Suresh S et al (2011) Aluminium-titanium diboride (Al-TiB_2) metal matrix composites: challenges and opportunities. Proc Eng 38:89–97. https://doi.org/10.1016/j.proeng.2012.06.013
5. Mohanavel V, Ravichandran M, Anandakrishnan V, Pramanik A et al (2021) Mechanical properties of titanium diboride particles reinforced aluminium alloy matrix composites: a comprehensive review. Adv Mater Sci Eng 18:7602160. https://doi.org/10.1155/2021/7602160
6. Sharma VK, Kumar V, Joshi RS (2019) Investigation of rare earth particulate on tribological and mechanical properties of Al-6061 alloy composites for aerospace application. J Market Res 8(4):3504–3516. https://doi.org/10.1016/j.jmrt.2019.06.025
7. Arora G, Sharma S (2021) Influence of rare earth addition on the properties of AA6351 hybrid composites. J Eng Res 9. https://doi.org/10.36909/jer.7901
8. Sahu M.K, Sahu R.K (2018) Fabrication of aluminum matrix composites by stir casting technique and stirring process parameters optimization. InTech. https://doi.org/10.5772/intechopen.73485
9. Liu K, Jiang X, Chen S, Yuan T, Yan Z (2022) Effect of SiC addition on microstructure and properties of Al-Mg alloy fabricated by powder and wire cold metal transfer process. J Market Res 17:310–319. https://doi.org/10.1016/j.jmrt.2022.01.014
10. Ellis M (2007) Joining of Al-based metal matrix composites - a review. Mater Manuf Processes 11:45–66. https://doi.org/10.1080/10426919608947460
11. Garcia R, Lopez VH, Bedolla E et al (2003) A comparative study of the MIG welding of AI/TiC composites using direct and indirect electric arc processes. J Mater Sci 38:2771–2779. https://doi.org/10.1023/A:1024471610718
12. Peng K et al (2011) Mechanical properties and wear resistance of aluminum composite welded by electron beam. Trans Nonferrous Metals Soc China 21(9):1925–1931. https://doi.org/10.1016/S1003-6326(11)60951-5
13. Ureña A et al (2000) Influence of interface reactions on fracture mechanisms in TIG arc-welded aluminium matrix composites. Compos Sci Technol 60(4):613–622. https://doi.org/10.1016/S0266-3538(99)00168-2
14. Yue TM, Xu JH, Man HC (1997) Pulsed Nd-YAG laser welding of a SiC particulate reinforced aluminium alloy composite. Appl Compos Mater 4:53–64. https://doi.org/10.1023/A:1008826513679
15. Kalaiselvan K, Dinaharan I, Murugan N (2020) Routes for the joining of metal matrix composite materials. Encyclopaedia Mater: Compos. https://doi.org/10.1016/B978-0-12-803581-8.11899-5
16. Devanathan C, Raja DE, Sonarn T, Ivanov M (2024) Effect of friction coefficient in friction stir welding of B_4C reinforced AA5083 metal matrix composites and use of fuzzy clustering technique for weld strength prediction. Adv Mater Sci Eng 1687–8434 https://doi.org/10.1155/2024/9880686
17. Ali KSA, Mohanavel V, Vendan SA et al (2021) Mechanical and microstructural characterization of friction stir welded SiC and B_4C reinforced aluminium alloy AA6061 metal matrix composites. Materials 14(11):3110. https://doi.org/10.3390/ma14113110
18. Abbasi M, Givi M, Ramazani A (2019) Friction stir vibration processing: a new method to improve the microstructure and mechanical properties of Al5052/SiC surface nanocomposite layer. Int J Adv Manuf Technol 100:1463–1473. https://doi.org/10.1007/s00170-018-2783-2
19. Barati M, Abbasi M, Abedini M (2019) The effects of friction stir processing and friction stir vibration processing on mechanical, wear and corrosion characteristics of Al6061/SiO_2 surface composite. J Manuf Process 45:491–497. https://doi.org/10.1016/j.jmapro.2019.07.034
20. Prater T (2014) Friction stir welding of metal matrix composites for use in aerospace structures. Acta Astronaut 93:366–373. https://doi.org/10.1016/j.actaastro.2013.07.023
21. Marie AF, Simar A (2016) A review about friction stir welding of metal matrix composites. Mater Charact 120:1–17. https://doi.org/10.1016/j.matchar.2016.07.010
22. Selvi S et al (2018) Cold metal transfer (CMT) technology - an overview. Defence Technol 14(1):28–44. https://doi.org/10.1016/j.dt.2017.08.002
23. Jicai F et al (2009) The CMT short-circuiting metal transfer process and its use in thin aluminium sheets welding. Mater Des 30(5):1850–1852. https://doi.org/10.1016/j.matdes.2008.07.015.4
24. Işıtan A, Aytekin S, Onar V (2024) Enhancing the microstructure and mechanical properties of cast aluminium matrix composites through nano-Al_2O_3 reinforcement via CMT and PMC welding. Inter Metalcast. https://doi.org/10.1007/s40962-023-01234-z
25. Pickin CG, Young K (2006) Evaluation of cold metal transfer (CMT) process for welding aluminium alloy. Sci Technol Weld Join. https://doi.org/10.1179/174329306X120886
26. Karimi MR, Wang SH, Jelovica J (2022) Taguchi-based experimental investigation into weld cladding of Ni-WC MMC overlays by CMT process. Int J Adv Manuf Technol 122:2433–2461. https://doi.org/10.1007/s00170-022-09816-7
27. Kesarwani S, Niranjan M, Singh V (2020) To study the effect of different reinforcements on various parameters in aluminium matrix composite during CNC turning. Compos Commun 22:100504. https://doi.org/10.1016/j.coco.2020.100504
28. Singh V, Niranjan MS, Murtaza Q (2024) Effect of Al-Si and Al-Mg filler wires on microstructure and mechanical properties of bead-on-plate weld on AA6061-T6 using CMT welding. Proc Inst Mech Eng Part E: J Process Mech Eng. https://doi.org/10.1177/09544089241241046
29. Verma RP, Pandey K, Sharma Y (2014) Effect of ER4043 and ER5356 filler wire on mechanical properties and microstructure of dissimilar aluminium alloys, 5083-O and 6061–T6 joint, welded by the metal inert gas welding. Proc Inst Mech Eng B: J Eng Manuf 229:1021–1028. https://doi.org/10.1177/0954405414535771
30. Elrefaey A, Ross NG (2015) Microstructure and mechanical properties of cold metal transfer welding similar and dissimilar aluminum alloys. Acta Metallurgica Sinica (English Letters) 28:715–724. https://doi.org/10.1007/s40195-015-0252-6
31. Huan P, Wang X, Zhang J (2020) Effect of wire composition on microstructure and properties of 6063 aluminium alloy hybrid synchronous pulse CMT welded joints. Mater Sci Eng A 790:139713. https://doi.org/10.1016/j.msea.2020.139713
32. Al-Sabur R, Slobodyan M, Chhalotre S, Verma M (2020) Contact resistance prediction of zirconium joints welded by small scale resistance spot welding using ANN and RSM models. Mater Today: Proc 47:5907–5911. https://doi.org/10.1016/j.matpr.2021.04.431

33. Al-Sabur R (2020) Tensile strength prediction of aluminium alloys welded by FSW using response surface methodology – comparative review. Mater Today: Proc 45:4504–4510. https://doi.org/10.1016/j.matpr.2020.12.1001
34. Sarabia L, Ortiz M (2010) Response surface methodology. Comprehensive Chemometrics: Chem Biochem Data Anal 1. https://doi.org/10.1016/B978-044452701-1.00083-1
35. Alam MA et al (2020) Modeling, optimization and performance evaluation of TiC/graphite reinforced Al 7075 hybrid composites using response surface methodology. Materials 14(16):4703. https://doi.org/10.3390/ma14164703
36. Karimi MR, Wang SH, Jelovica J (2023) Characterization of cold metal transfer and conventional short-circuit gas metal arc welding processes for depositing tungsten carbide-reinforced metal matrix composite overlays. Int J Adv Manuf Technol 128:2551–2570. https://doi.org/10.1007/s00170-023-11955-4
37. Anand R, Padmanabhan R (2023) Influence of friction stir welding process parameters and statistical behaviour of the novel interlock aluminium alloys joint with SiCp reinforcement.". CIRP J Manuf Sci Technol 45:260–270. https://doi.org/10.1016/j.cirpj.2023.07.002
38. Zhiqiang Z et al (2023) Optimization of porosity and surface roughness of CMT-P wire arc additive manufacturing of AA2024 using response surface methodology and NSGA-II. J Market Res 24:6923–6941. https://doi.org/10.1016/j.jmrt.2023.04.259
39. Koli Y, Yuvaraj N, Vipin AS (2020) Multi-response mathematical modeling for prediction of weld bead geometry of AA6061-T6 using response surface methodology. Trans Indian Inst Met 73:645–666. https://doi.org/10.1007/s12666-020-01883-2
40. Wang B, Xue S, Ma C, Wang J, Lin Z (2017) Effects of porosity, heat input and post-weld heat treatment on the microstructure and mechanical properties of TIG welded joints of AA6082-T6. Metals 7(11):463. https://doi.org/10.3390/met7110463
41. Wang L et al (2017) Effects of welding parameters on microstructures and mechanical properties of disk laser beam welded 2A14-T6 aluminium alloy joint. J Manuf Process 31:240–246. https://doi.org/10.1016/j.jmapro.2017.11.017
42. Zhao Y, Chen F, Cao S, Chen C, Xie R (2022) Effect of CMT welding heat input on microstructure and properties of 2A14 aluminum alloy joint. Metals 12(12):2100. https://doi.org/10.3390/met12122100
43. Tiryakioğlu M (2020) The effect of hydrogen on pore formation in aluminium alloy castings: myth versus reality. Metals 10(3):368. https://doi.org/10.3390/met10030368
44. Tiryakioğlu M (2020) Solubility of hydrogen in liquid aluminium: reanalysis of available data. Int J Cast Met Res. https://doi.org/10.1080/13640461.2020.1718337
45. Gupta SK et al (2018) Effect of welding speed on weld bead geometry and percentage dilution in gas metal arc welding of SS409L. Mater Today: Proc 18:5032–5039. https://doi.org/10.1016/j.matpr.2019.07.497
46. Legait P (2006) Formation and distribution of porosity in Al-Si welds. Worcester Polytechnic Institute (PhD Thesis)
47. Mvola B, Kah P (2017) Effects of shielding gas control: welded joint properties in GMAW process optimization. Int J Adv Manuf Technol 88:2369–2387. https://doi.org/10.1007/s00170-016-8936-2
48. Schempp P, Rethmeier M (2015) Understanding grain refinement in aluminium welding. Weld World 59:767–784. https://doi.org/10.1007/s40194-015-0251-2
49. Kumar NP et al (2016) Investigations on the parametric effects of cold metal transfer process on the microstructural aspects in AA6061. J Alloy Compd 658:255–264. https://doi.org/10.1016/j.jallcom.2015.10.166
50. Schaffer P et al (2006) Segregation of particles and its influence on the morphology of the eutectic silicon phase in Al-7 wt.% Si alloys. Scripta Mater 54. https://doi.org/10.1016/j.scriptamat.2005.10.016
51. Chidambaram A et al (2019) Microstructure and mechanical properties of AA6061–5wt. %TiB_2 in-situ metal matrix composite subjected to equal channel angular pressing. Mater Sci Eng: A 759762–769. https://doi.org/10.1016/j.msea.2019.05.068
52. Sun C, Lu N, Liu H et al (2020) Effect of necklace-type distribution of SiC particles on dry sliding wear behavior of As-cast AZ91D/SiCp composites. Crystals 10(4):296. https://doi.org/10.3390/cryst10040296
53. Shang J et al (2013) Microstructure characteristics and properties of Mg/Al dissimilar metals made by cold metal transfer welding with ER4043 filler metal. Rare Metal Mater Eng 42(7):1337–1341. https://doi.org/10.1016/S1875-5372(13)60081-8
54. Lakshminarayanan AK, Balasubramanian V, Elangovan K (2009) Effect of welding processes on tensile properties of AA6061 aluminium alloy joints. Int J Adv Manuf Technol 40:286–296. https://doi.org/10.1007/s00170-007-1325-0
55. Wei N, Cui HZ, Wu J et al (2015) Effects of forming conditions and TiC–TiB_2 contents on the microstructures of self-propagating high-temperature synthesized NiAl–TiC–TiB_2 composites. Acta Metall Sin (Engl Lett) 28:39–47. https://doi.org/10.1007/s40195-014-0163-y
56. Hansen N (2004) Hall-Petch relation and boundary strengthening. Scripta Mater 51:801–806. https://doi.org/10.1016/j.scriptamat.2004.06.002
57. Hakem M, Lebaili S, Mathieu S et al (2019) Effect of microstructure and precipitation phenomena on the mechanical behavior of AA6061-T6 aluminium alloy weld. Int J Adv Manuf Technol 102:2907–2918. https://doi.org/10.1007/s00170-019-03401-1
58. Tingting Z et al (2020) Effects of rare earth elements on the microstructure and wear properties of TiB_2 reinforced aluminium matrix composite coatings: experiments and first principles calculations. Appl Surf Sci 530:147051. https://doi.org/10.1016/j.apsusc.2020.147051
59. Venkatesh VSS et al (2023) Processing and evaluation of nano SiC reinforced aluminium composite synthesized through ultrasonically assisted stir casting process. J Market Res 24:7394–7408. https://doi.org/10.1016/j.jmrt.2023.05.030
60. Pazhouhanfar Y, Eghbali B (2018) Microstructural characterization and mechanical properties of TiB2 reinforced Al6061 matrix composites produced using stir casting process. Mater Sci Eng A 710:172–180. https://doi.org/10.1016/j.msea.2017.10.087
61. Jing X et al (2012) Behavior of CeO_2 additive in in-situ TiB_2 particles reinforced 2014 Al alloy composite. Trans Nonferrous Metals Soc China 22(5):1012–1017. https://doi.org/10.1016/S1003-6326(11)61277-6
62. Srivatsan TS, Prakash A (1994) The quasi-static fracture behavior of an aluminium alloy metal-matrix composite. Compos Sci Technol 54(3):307–315. https://doi.org/10.1016/0266-3538(95)00057-7
63. Wang J et al (2022) Effect of preheat & post-weld heat treatment on the microstructure and mechanical properties of 6061–T6 aluminium alloy welded sheets. Mater Sci Eng, A 841:143081. https://doi.org/10.1016/j.msea.2022.143081
64. Ahmad R, Bakar MA (2011) Effect of a post-weld heat treatment on the mechanical and microstructure properties of AA6061 joints welded by the gas metal arc welding cold metal transfer method. Mater Des 32(10):5120–5126. https://doi.org/10.1016/j.matdes.2011.06.007
65. Rajendran C et al (2024) Enhancing tensile properties of pulsed CMT–MIG welded high strength AA2014-T6 alloy joints: effect of post weld heat treatment. Int J Lightweight Mater Manuf 7(2):344–352. https://doi.org/10.1016/j.ijlmm.2023.10.004

66. Li X, Yan H, Wang ZW et al (2019) Effect of heat treatment on the microstructure and mechanical properties of a composite made of Al-Si-Cu-Mg aluminium alloy reinforced with SiC particles. Metals 9(11):1205. https://doi.org/10.3390/met9111205

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

MĀPAN-Journal of Metrology Society of India
https://doi.org/10.1007/s12647-024-00764-z

ORIGINAL PAPER

Evaluation of Tribological and Cooling Performance of TiN and DLC-Coated Pistons for Miniature Stirling Cryocooler

M. Singhal[1,2]*, R. Kumar[1], R. S. Walia[3] and S. K. Pandey[4]

[1]Department of Mechanical Engineering, Delhi Technological University, Shahbad Daulatpur, Bawana Road, Delhi 110042, India

[2]Defence Research and Development Organisation (DRDO) SSPL Complex, Timarpur, Delhi 110054, India

[3]Department of Production and Industrial Engineering, Punjab Engineering College, Sector 12, Chandigarh 160012, India

[4]Defence Research and Development Organisation (DRDO), Metcalfe House, Delhi 110054, India

Received: 03 March 2024 / Accepted: 13 July 2024

© Metrology Society of India 2024

Abstract: The Stirling cryocoolers are widely used in the military and aerospace fields due to advantages such as high efficiency and compact structure. The pistons of these coolers with very small clearances operate at high speed and without lubrication. Piston failure in the coolers has often affected the operational life of the cooler. The objective of the study was to find a solution to piston failure by providing a surface coat to the pistons, carrying out their tribological evaluation, and analyse cooler thermal performance to identify the better of two coatings for field applications. In the present experimental study, a Stirling cooler developed with a cooling capacity of 0.5 W @80 K with Helium as the working fluid was integrated with surface-coated pistons and tested for its mechanical and thermal performance. On testing, DLC coating exhibits better wear resistance than TiN coating and therefore is a more suitable coating for cryocooler pistons. The present study clearly shows that DLC-coated pistons reduce power consumption and increase the number of operational hours of miniature Stirling cryocoolers.

Keywords: Stirling; Cryocooler; Titanium nitride (TiN); Diamond-like Carbon (DLC); Infrared

1. Introduction

Research on miniature mechanical cryocoolers with investigation of different types of refrigeration cycles for small cryocoolers has continued since the earliest interest in infrared sensors cooling in thermal cameras, space exploration, earth observation, and medicines (MRI). Stirling cryocoolers are particularly attractive for miniature units because of their compact configuration, low weight, and high efficiency in terms of power input per unit of refrigeration produced. They are designed to be unlubricated and wear-free, to meet the requirements of high reliability, high efficiency, and long maintenance-free life.

All dynamic machines are expected to meet the requirement of good mechanical as well as tribological performance over a wide and changing set of environmental conditions. Surface coating continues to be the most prevalent and cost-effective method in reducing the tribological issues related to friction and wear. The compressor piston in a cryocooler reciprocates at around 75 Hz (stroke of 2 mm) inside the cylinder in a Helium atmosphere. It compresses Helium to around 30 bars with a pressure ratio of around 1.2. The working of the piston under high pressure at very high speed with very small clearances increases wear possibility. This makes the cooler susceptible to early failure as both cylinder and piston are of the same material and hardness, and are also ground. To overcome this, the piston needs to have sufficient hardness at the surface and a low coefficient of friction between the piston-cylinder assembly. To fulfill this requirement, the piston is coated with harder surface coatings in comparison to the cylinder. In addition, the dimensional aspects of the piston are maintained with low coating temperature and submicron thickness uniformity. Ceramic and carbon coatings such as titanium nitride (TiN)

*Corresponding author, E-mail: mayanksinghal25@gmail.com

Published online: 24 July 2024

and diamond-like carbon (DLC) coatings possess good mechanical properties and chemical inertia. Various conventional binary coatings such as TiN, CrN, TiC, etc., are used for improving tribological behaviour between moving and contacting surfaces like piston cylinder liner assembly, bearings, workpieces, and tools.

In the present case, Titanium Nitride and Diamond Like coating are the most suitable candidates as they are harder than hardenable steels. Titanium nitride (TiN) is a metal nitride hard ceramic coating having remarkable tribo-mechanical properties like high hardness, low rate of wear and corrosion [1–5] with a broad range of tribological applications in machine elements, turbine bearings, aerospace, automotive, and cutting tools. Extensive research on the areas of application of TiN coatings, their characterization, and deposition technique has been carried out to highlight the advantages of TiN coatings over uncoated surfaces of the mechanical components and tools.

Aytacet et al. [6] studied the wear of TiN-coated AISI 3343 steel used extensively in the moulding sector and Ghani et al. [7] investigated the wear mechanism of TiN-coated carbide and uncoated cermet's tools for end milling of AISI H13 tool steel by varying combinations of feed rate, depth of cut, and speed of cutting. Koeskiet et al. [8] compared CVD- and PVD-deposited TiN coatings for their cutting performance and wear properties during the continuous cutting of Ni-based super alloys. In a study by Xian et al. [9], TiN coating was deposited on three different tool materials i.e., high-speed steel, cemented carbide, and TiCN-based cermet using arc evaporation. The effect of tool materials on the microstructure, composition, hardness, adhesion, and tribological properties of the coatings was investigated to explore the relationship between the substrate material and the microstructure and mechanical properties of the coating. Hsu et al. [10] evaluated the performance of titanium nitride (TiN) coated Fiber Bragg grating (FBG) sensors in cryogenic temperature environments. Single metal-coated and bare FBG sensors were placed in an aluminium alloy mould, packaged in a thermally insulated, and used as cryogenic temperature sensors. TiN coating on FBG sensors changed nonlinear characteristics of the thermo-optic and thermal expansion coefficients resulting in accurate cryogenic temperature measurements. In addition to TiN coatings, Diamond-like carbon (DLC) coating is also widely used because of its good tribological characteristics and aesthetic value. DLC coating on machine parts allows them to operate under heavy loads, high temperatures ($\sim$ 500 °C), and corrosive environments. Diamond-like carbon (DLC) coatings are also relatively popular due to their superior properties like high hardness and lower coefficient of friction.

Li et al., [11] carry out a progress review and discuss various coatings, such as Cr-Al_2O_3 coating prepared by a process of electroplating, CrN and DLC coatings prepared by physical vapor deposition (PVD), and Mo-based and C-based coatings prepared by the thermal spraying process. In addition, development trends of commonly used piston rings are summarised and the pros and cons of commonly used piston ring coatings are also discussed.

DLC coatings find extensive application in the automotive industry [12–14] and are also applied to sliding gasoline engine parts. DLC coatings effectively reduce friction under an electrohydrodynamic lubrication (EHL) region [15] and are thus used to coat piston rings and valve lifters. Various studies have also been done on other DLC-coated mechanical components such as bearings [16], gears [17], and cutting and forming tools [18, 19]. Hanzelka et al. [20] proposed a DLC layer as an anti-seizing coating for their cryogenic apparatus and carried out measurements of thermal radiative properties such as thermal emissivity and absorptivity in the range of cryogenic and room temperature.

In certain applications of cryogenic cooling, a temperature down to 80 K is required, and cooling power varies from milli-watt to a few Watts. One such defence application is the cooling of infrared (IR) sensors to cryogenic temperature to minimise background 'noise' and get a sharp thermal image [21]. A reliable refrigeration system is necessary for cooling the detector to low temperatures and thus reducing the noise. In this application, isolating the sensor from mechanical shock, vibration, and electromagnetic interference is also important. In addition to the Joule Thomson (JT) cryocooler, a miniature cryocooler working with the reversed Stirling cycle principle is a solution for cooling such detectors. Miniature Stirling cryocoolers are widely used in an infrared detector assembly for maintaining cryogenic temperature. For optimum performance, the life of the Stirling cryocooler is a critical parameter. Several investigations [22, 23] have been carried out to address the cryocooler modelling, optimization of the design parameters [24, 25], and regenerator design of the Stirling cycle cryocooler [26–29]. However, limited literature is available on the cryocooler compressor piston design [30] and no studies are available on the effect of coatings on the thermal performance and operability of the cryocooler with a coated piston.

Stirling cryocooler compressor pistons with small clearances function under high pressure and speeds in the absence of lubrication increasing the wear possibility. Piston seizure and failure were identified to be one of the key reasons for cooler failure in field operations. The motivation for carrying out this study was to eliminate the possibility of cooler failure due to failure of piston by providing a hard coat and increasing the operational life of the coolers in field applications.

In the present investigation, the effect of TiN and DLC coating on the thermal performance and operational life of the Stirling cooler has been studied with two primary objectives. The first objective is to determine a suitable surface coating method and material for miniature Stirling cryocooler pistons, while the second objective of the present work is to carry out mechanical testing and thermal performance testing of a miniature Stirling cryocooler integrated with coated pistons on an in-house created experimental test setup. In this regard, the tribological performance of pistons with TiN and DLC coating was evaluated.

2. Experimental Study

2.1. Coating Methodology

A semi-industrial sputtering system was used for the deposition of TiN coatings. The effective plasma volume of the sputtering system was 0.005832 m^3. The sputtering system mainly consists of a cylindrical double-walled water-cooled vacuum chamber with several ports, which act as the sputtering chamber. It houses the magnetron cathodes, the substrate holder, feed-throughs, etc. A high vacuum pumping system with a turbomolecular pump is used to create a vacuum in the chamber of the order of 5.6×10^{-6} mbar. The coatings were deposited at a substrate temperature in the range of 250–325 °C. DLC coating was added to the piston surface by the Plasma-assisted Chemical Vapour Deposition (PACVD) method by magnetron sputtering techniques. The processing temperature (< 250 °C) at which the material properties of the parent material remain unaffected. The general coating parameters for deposition of TiN and DLC coatings on the pistons are listed in Table 1.

A piston cylinder is shown in Fig. 1. The required thickness of the coating on the piston is within 2–3 µm and the uniformity of thickness within 1 µm is a crucial requirement for the maintenance of the part geometry. The parent material before the coating is M42 alloy steel, hardened and ground to a hardness of 55–60 HRC with a surface roughness value of Ra ~ 0.20 µm. After coating, the piston-cylinder pair is lapped and polished to the surface finishes of the order of 0.05 µm Ra. TiN-coated and DLC-coated piston samples are shown in Fig. 2.

Table 1 General parameters for TiN and DLC coating

Parameters	TiN Coating	DLC Coating
Coating technique	PVD	PACVD
Colour of coating	Golden Brown	Black
Coating temperature (°C)	250–325	< 250
Friction coefficient	0.4	0.1
Micro hardness HV 0.5	691	1173
Thermal conductivity (W/mK)	19.2	500

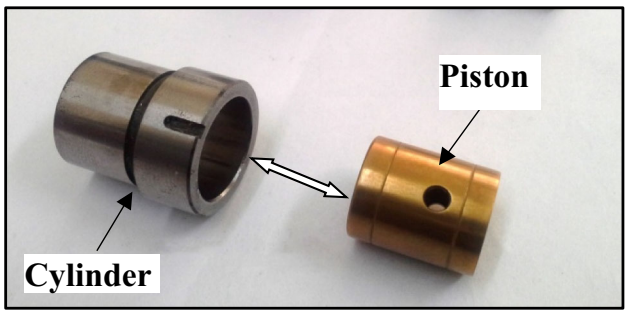


Fig. 1 Piston and cylinder pair of a stirling cryocooler

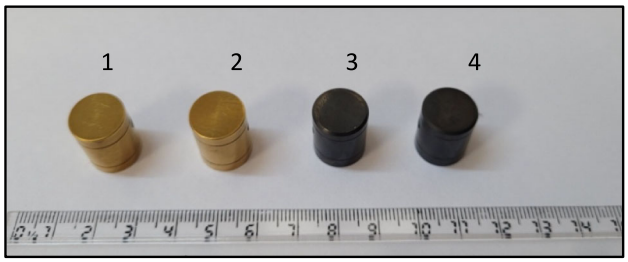


Fig. 2 Piston samples (1&2—TiN coated, 3&4—DLC coated)

2.2. Characterization of the Coating

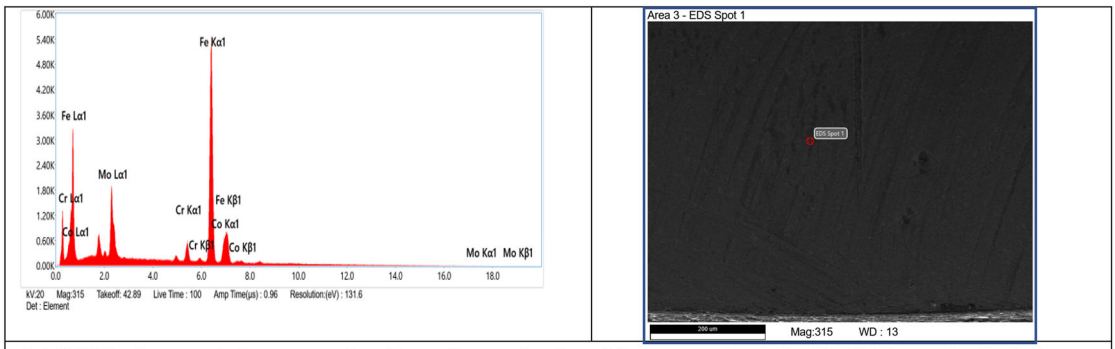
The SEM and EDS results of the coated samples are shown in Fig. 3. While the SEM micrographs reveal the morphology of the coatings, the chemical composition of the coating was evaluated using EDS. The energy spectrum as presented alongside the micrographs, clearly reveals the peaks of titanium in the coated sample and the peak of carbon in the DLC-coated sample are visible in the test results. The elemental concentration of TiN and DLC coating were identified by a scanning microscope coupled with energy-dispersive X-ray spectroscopy (EDS). EDS

Table 2 Elemental concentration of TiN sample

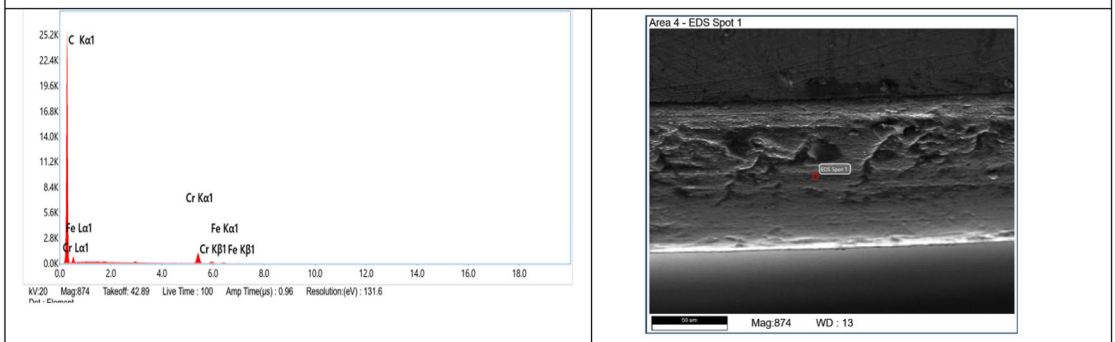
Element	N	Ti	Fe
Weight %	19.83	77.56	2.62

Table 3 Elemental concentration of DLC sample

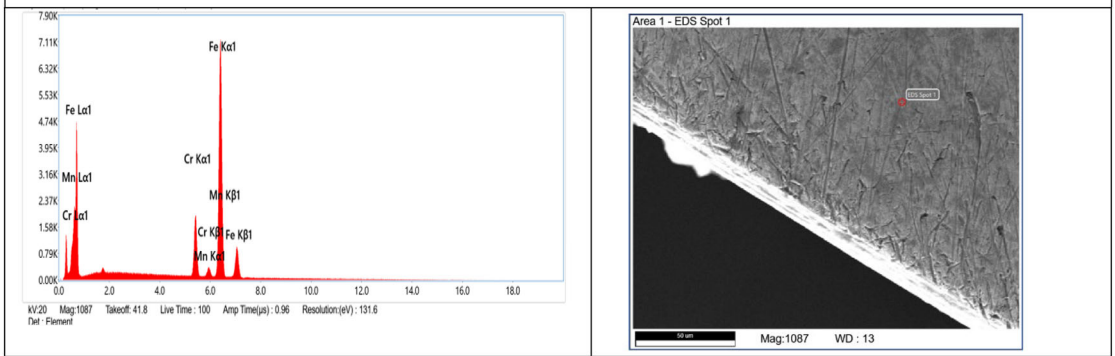
Element	C	Cr	Ar
Weight %	74.28	24.57	0.75



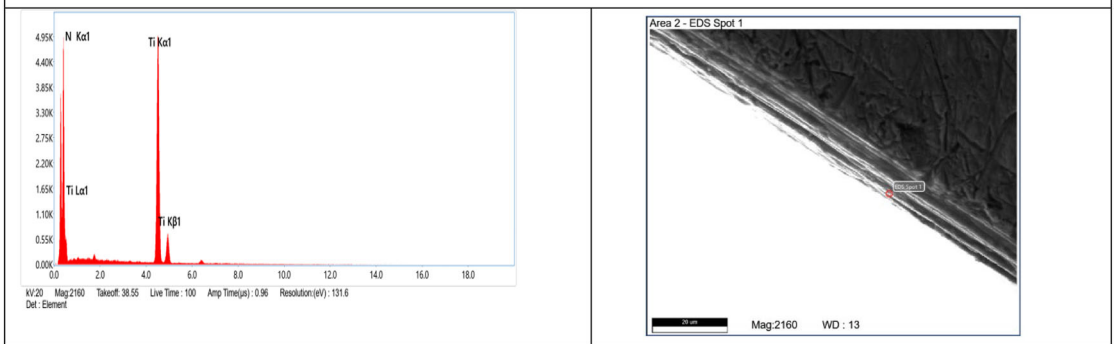
(a) EDS elemental concentration & analysis of base material of DLC-coated piston



(b) EDS elemental concentration & analysis of DLC coating on the piston



(c) EDS elemental concentration & analysis of base material of TiN-coated piston



(d) EDS elemental concentration & analysis of TiN coating on the piston

Fig. 3 EDAX elemental concentration & analysis of base material and both coatings

Table 4 Microhardness of TiN and DLC coated piston on HV0.5 scale

Sample	Mean	Standard deviation
TiN coated	691.92	15.81
DLC coated	1173.29	78.13

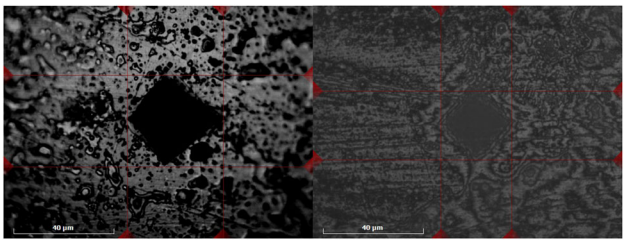


Fig. 4 Microstructure of TiN and DLC samples after indentation (left to right)

analysis with the elemental concentration corresponding to each coating is shown in Tables 2 and 3.

The microhardness measurements of the samples were performed using the DURAMIN-40 A2 testing machine. The testing method was VICKERS on HV 0.5 scale and the dwell time was 15 s. The surface hardness on the TiN and DLC-coated samples is given in Table 4. The readings are taken at four different points to consider the repeatability.

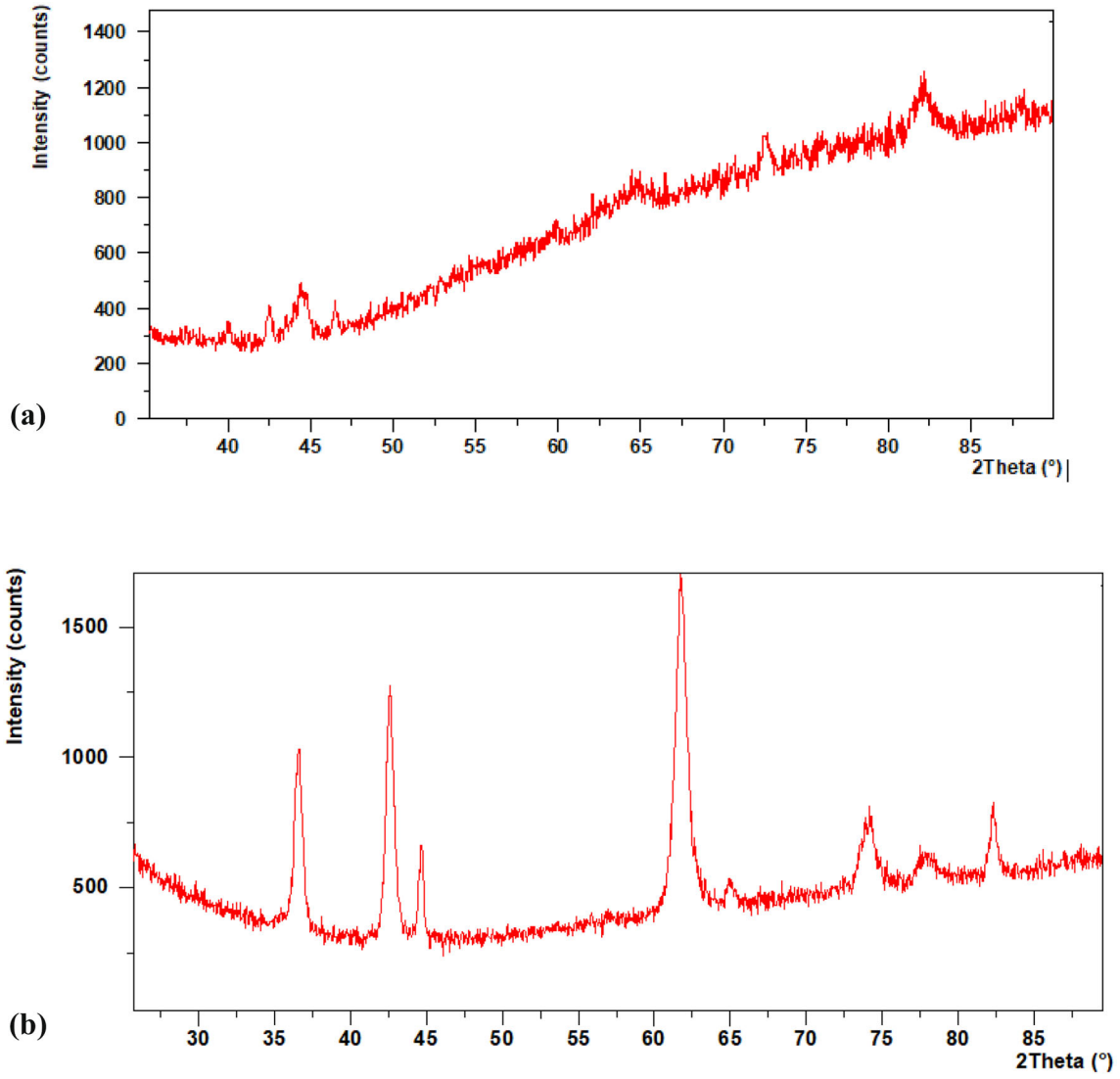


Fig. 5 **a** Grazing incidence X-ray diffraction (GIXRD) patterns of DLC-coated Piston. **b** Grazing incidence X-ray diffraction (GIXRD) patterns of TiN-coated Piston

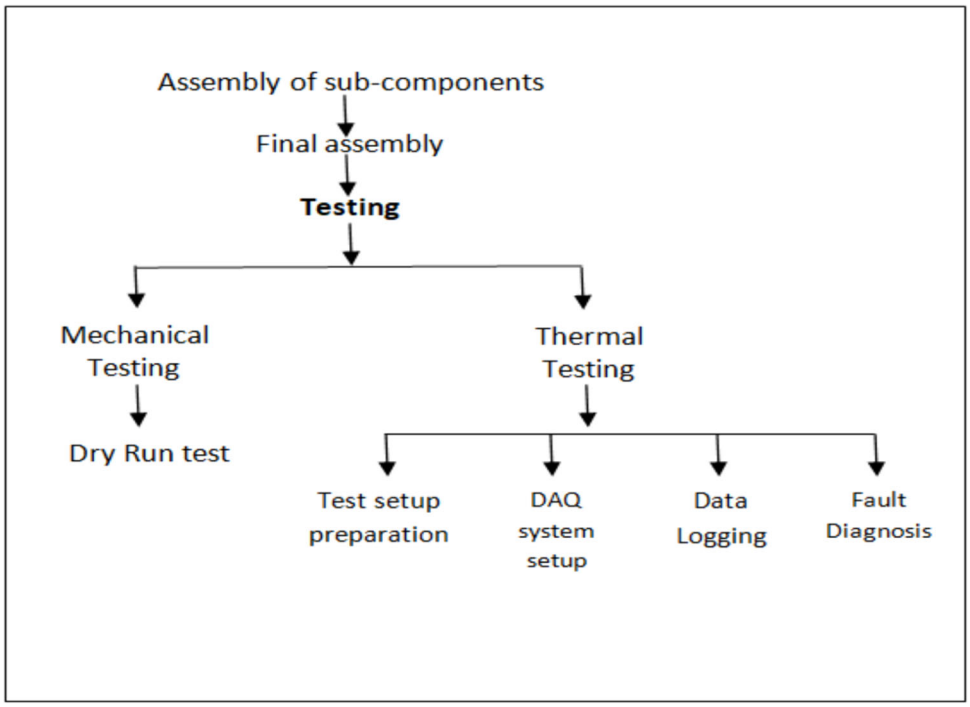


Fig. 6 Experimental testing procedure

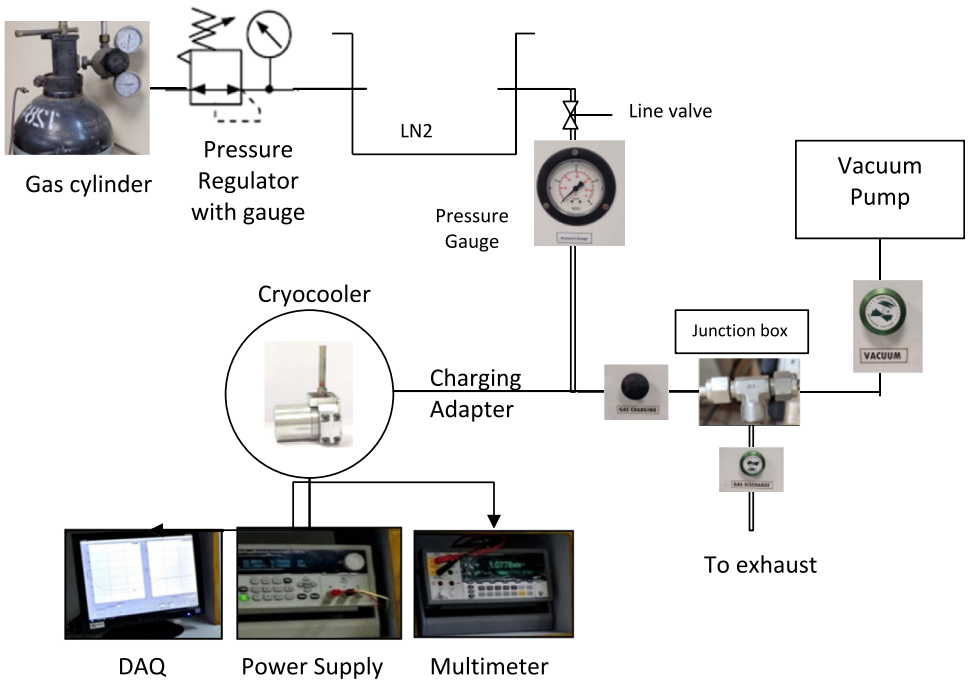


Fig. 7 Laboratory set up for prototype cryocooler testing

Table 5 The test requirements for a cryocooler

Cold tip temperature	80 K ± 1 K
Cool down time	< 8 min
Max input power	< 15 Watt
Steady-state power input	< 10 Watt
Room temperature	25 ± 2 °C

Figure 4 shows the indentation imprints by diamond indenter on the samples after 500 gf normal load was applied during the hardness test. The shape of the indentation marks reveals a pile-up in TiN-coated samples, while in the case of DLC-coated samples, no such pile-up was observed. The reason for the same is the softer coating.

Further, GIXRD measurements were also carried out on the samples. The pattern developed in DLC and TiN-coated

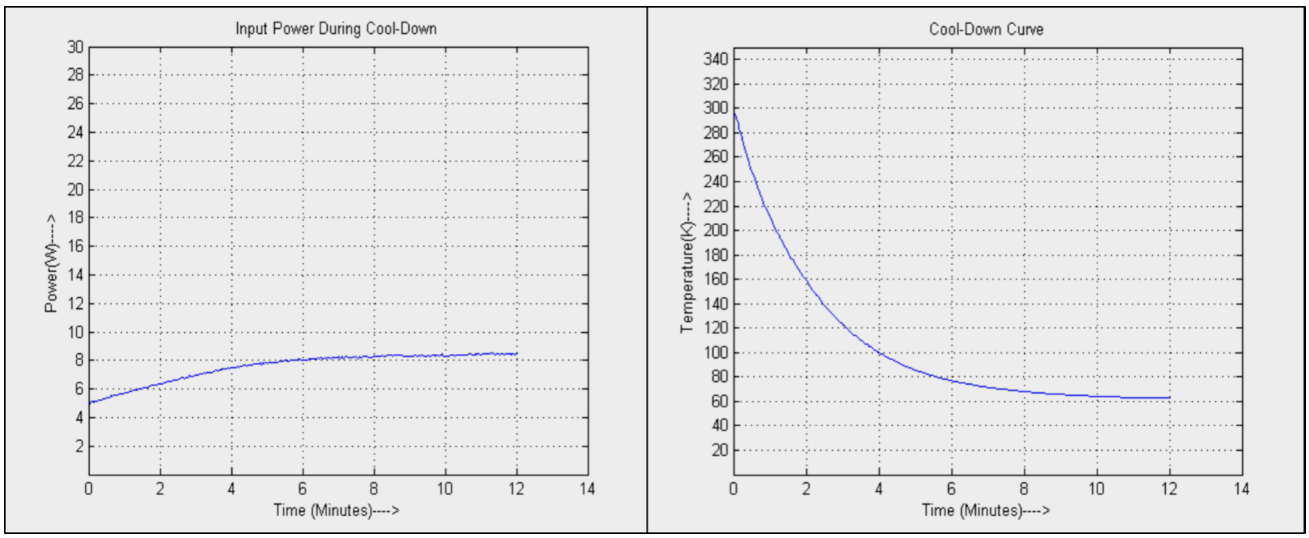


Fig. 8 Screen shot of data acquisition system of a typical performance run

samples is shown in Fig. 5a, b respectively. Phase analysis of the coated samples reveals that the most intense diffraction peaks originate mainly due to the Ti microparticles on the coating surface. On the other hand, the DLC-coated sample shows the presence of the graphite phase.

2.3. Testing of the Coated System

The cryocooler is tested on a dedicated experimental rig where the output parameters are continuously monitored through a dedicated data acquisition system. The data acquisition system generates an automated report at the end of every experiment containing the experimental results and system configuration. The cryocooler is charged and purged several times (10–12 times) to eliminate contaminations in the working medium like moisture, dirt, grease, etc. For the purging process, the cryocooler is coupled with a rotary vacuum pump. This process is carried out at room temperature and the cryocooler is not in operative mode. The helium gas charging line also includes a cryogenic trap in which the coil is immersed in liquid nitrogen to the filling station and a filter to remove the moisture/contaminations present in the gas. After the repetitive purging process, finally, the cryocooler is refilled with pure (purity 99.99%) and fresh Helium gas at 26–28 bar charge pressure.

After the final assembly, the prototype cryocooler is to be tested for its mechanical and thermal performance as shown in Fig. 6. In the mechanical testing, the dry run without the charging gas is carried out for several hours to monitor the abnormal sound, vibration, and power characteristics. The prototype undergoes thermal testing with helium gas charged if it passes through the dry run test. In thermal testing, the cryocooler is sealed through various stages and finally with a charging screw placed to ensure any type of gas leakage. A helium leak detector in sniffer mode is used to check the gas leakage from the outer casing of the main body of the cryocooler. The permissible value of the rate of cooler filling gas leakage is less than 4×10^{-8} atm cc/sec. After successfully passing this test, a RTD (Pt-100) temperature sensor is mounted at the cold end of the test Dewar to monitor the temperature available at the cold end. In the thermal performance, the cooldown time, the temperature, and the power input to the cooler are monitored. Continuous data logging is performed by a dedicated data acquisition system. The cold-finger region, on which the RTD is mounted, is maintained in a vacuum of the order of 10^{-7} mbar with the help of a turbo vacuum pump (Pfeiffer HiCube 80), to prevent any formation of frost due to cryogenic temperatures, which eventually may cause damage or may result in measurement variations. A programmable DC power supply (Tektronix PWS4305, 0–30 V) is used to supply and record the input electrical parameters like current, voltage, etc. A multimeter (Fluke 8845A) is used to monitor the temperature parameters and a Line valve (Nupro make), and a Pressure gauge (Wika make Range 0–40 kg/cm^2) are other components of the setup.

The complete lab setup to test the cryocooler prototype is shown in Fig. 7. Initially, the power input to the cooler is high until the system has cooled to its operating temperature (80 K) which is about 4–6 min. After achieving that condition, the first few hours of operation may see some further reduction in the input power in terms of a reduced motor supply voltage and hence machine speed. During the thermal test, if there is a flaw in desired performance parameters, the cooler undergoes extensive fault diagnosis

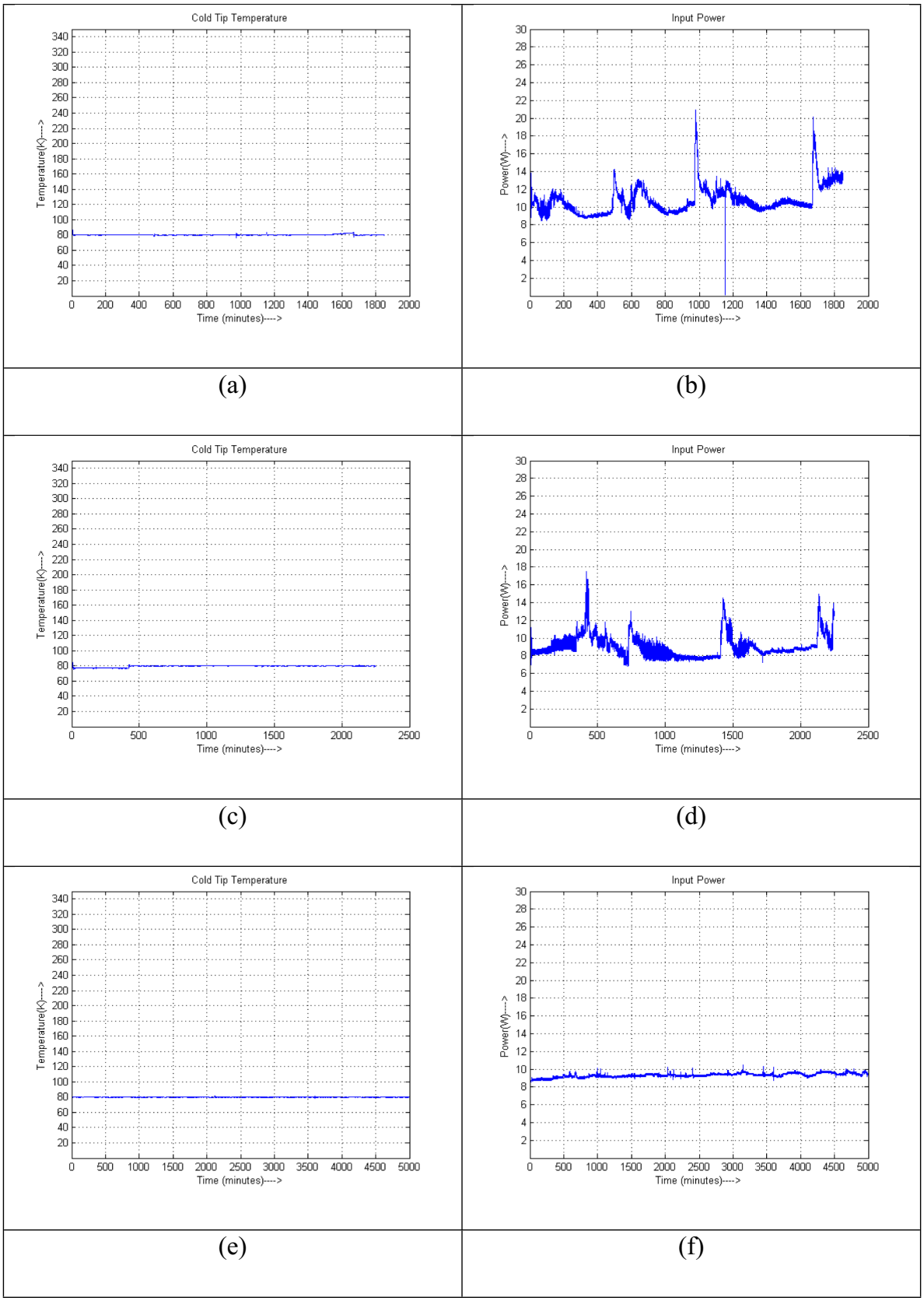
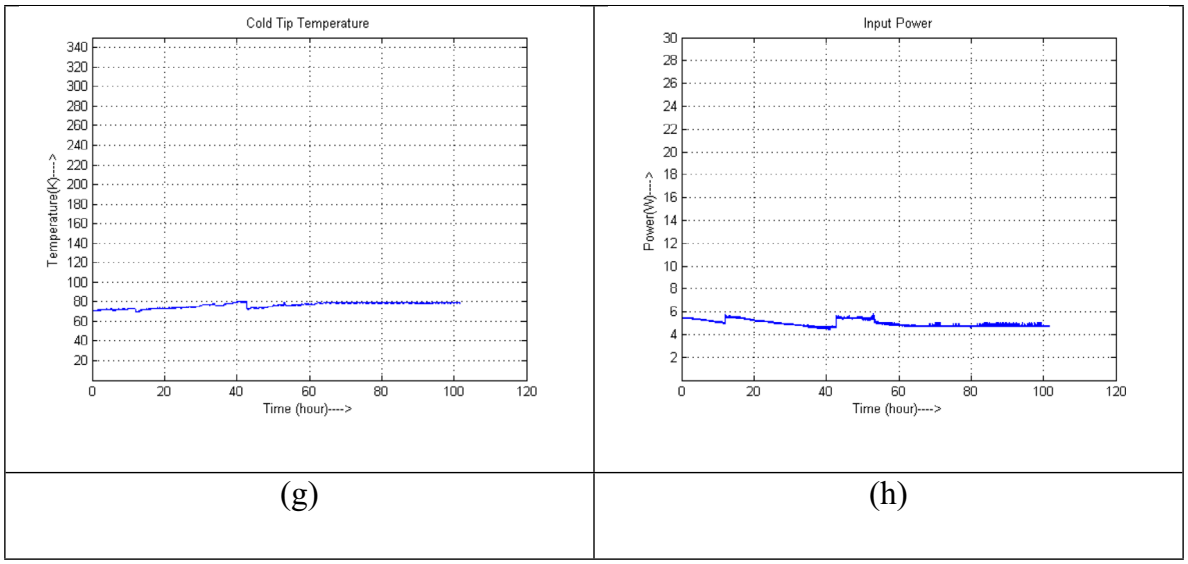


Fig. 9 Time vs cold tip temperature (left) and power characteristics (right) for TiN coated piston **a**–**d** & DLC coated piston **e**–**h**



(g) (h)

Fig. 9 continued

for a series of temperature and power checks at various levels.

3. Results and Discussions

To investigate the effect of TiN and DLC coating on cryocoolers, four-piston samples (two each) have been tested with a cryocooler. Initially, two coolers were sealed and assembled with TiN-coated pistons after charging the refrigerant (He) gas. A dummy dewar was used to create a vacuum in the range of 10^{-7} mbar around the cold tip of the cryocooler with a Turbo molecular pump (TMP). A calibrated RTD (PT-100) was used for cold tip temperature measurement. A data acquisition system was put in place for monitoring and logging the data. The testing requirement criteria are given in Table 5 for these experiments.

Results of a prototype cryocooler that met the performance criteria in a typical test run were obtained with the help of a dedicated data acquisition system. The results in terms of power characteristics and cold tip temperature of the cryocooler are shown in Fig. 8.

The cryocooler parameters maintained for testing the piston and cylinder assembly in the different prototype cryocoolers are given in Table 6. The Experimental test results in terms of run hours, cold tip temperature, cool down time & power drawn by the coolers are shown in Table 7.

The results obtained in terms of power characteristics and cold tip temperature are shown in Fig. 9. The four plots (cool down temperature & power vs time) are shown for two sets of each piston coating TiN and followed by DLC

Table 6 Cryocooler parameters during the experiment

Nominal speed	4200 RPM
Charging medium	Helium gas
Charging pressure	28 bar
Stroke length	2.2 mm
Piston diameter	14 mm
Ambient temperature	300 K
Cold tip temperature	80 K

Table 7 Test results obtained during the experimentation

Cooler parameter	Cooler-1 TiN coated	Cooler-2 TiN coated	Cooler-1 DLC coated	Cooler-2 DLC coated
Total duration of run	30 h 9 min	37 h 29 min	83 h 11 min	101 h 40 min
Cold tip temperature	76.10 K	75.40 K	78.30 K	69.90 K
Cool down time	4 min 47 s	6 min 25 s	6 min 47 s	4 min 05 s
Max power input	14.15 W	11.30 W	11.0 W	5.79 W
Max Power input throughout the experiment	20.90 W	17.54 W	11.0 W	5.72 W

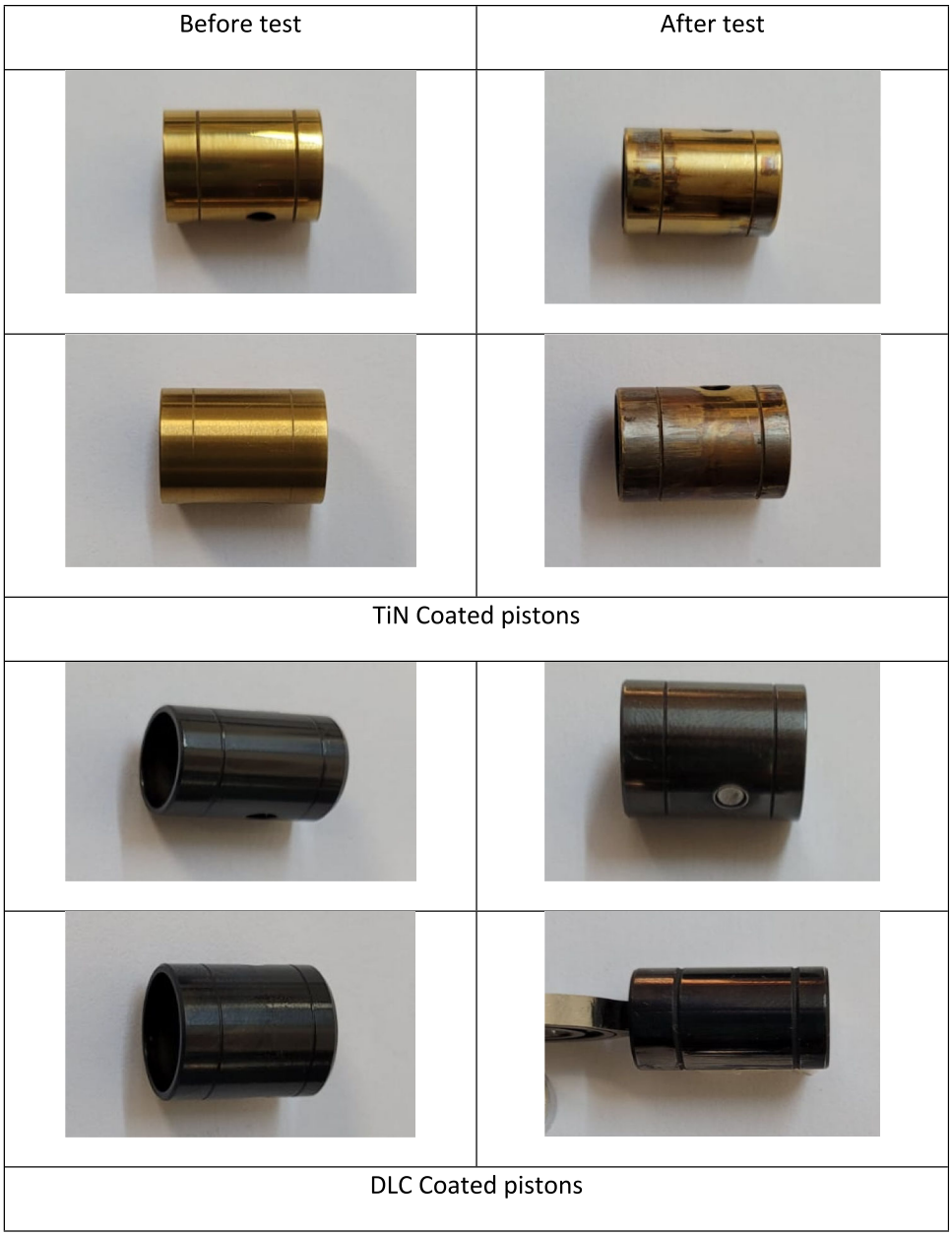


Fig. 10 TiN & DLC coated pistons before and after the cooler operation

coated with the parameters shown in the table at two prototype cryocoolers.

The other parameters of the test results such as time duration of the test run, cold tip temperature during the test, cool down time & power drawn by the coolers are tabulated in Table 7 below:

The cooler-1 was clocked around 30 h continuously. Initially, for the first few hours, it operated without an increase in power. The temperature during this period was also around 77 K, less than the set value of 80 K. The audible noise observed during this period was also negligible (permissible level < 35 dB). However, after a few hours, the power gradually rose till 30 h of run and its peak value went up to more than 20 Watts along with a substantial increase in audible sound levels. A cyclic rise and fall in power levels were seen and once power reached a certain high level, a high audible sound from the cooler was observed. Due to the continuous high power being drawn and a very high audible sound from the cooler the testing had to be discontinued. A similar pattern was observed for cooler 2, during 37 h of continuous run. Figure 9b–d shows the power vs time and Fig. 9a–c shows the cold tip temperatures vs time logged on the data acquisition system. Both the coolers failed to meet the required thermal test specifications. The cooler was disassembled to see the reasons for failure, especially for the power rise. The following were the main observations:

a) A brown powder-like substance could be seen on the components such as the piston & cylinder, crankcase, displacer assembly, etc. Brownish color powder could also be seen on the inner side of the main body
b) Several rubbing marks on the piston and sleeve assembly were seen and this hindered optimal operation since clearance space was filled with powder/debris.
c) The rest of the components showed no changes.

The images of the piston before and after assembly are shown in Fig. 10. There were visible rubbing and scratch marks on the piston/cylinder showed some signs of drawing a high current. This is looking to be the most probable cause of this failure as the coating comes out in the form of brown powder. Both the coolers were opened and reassembled with fresh DLC-coated pistons. Cooler-1 operated continuously for 83 h and cooler-2 operated for 101 h without any sign of a mechanical seizure, abnormal sound, vibration, or any surge in current/power during the period of testing. The power drawn and cold tip temperature measured during the experiment are shown in Fig. 9e–h. All the specified thermal test specifications were met successfully. The testing was stopped and similar tests were carried out after regular periods of intervals with satisfactory results.

4. Conclusion

In this study, a Stirling cryocooler with titanium nitride (TiN) and Diamond-like coating (DLC) pistons has been investigated experimentally for its thermal and operational performance involving measurement of its cooling, power characteristics, and the number of operational hours. The following major conclusions can be drawn from experimentations:

(i) The TiN-coated pistons showed less adhesion and more abrasion during cooler operation. The primary cause of brown powder observed in most components was the wear of coating between the piston (compressor) and the sleeve. DLC coating exhibited minimum or no visual transference of wear debris from the DLC coating during testing.
(ii) The TiN-coated piston samples could not hold for more than 40 h cycle whereas DLC-coated piston samples sustained more than 100 h cycles of cryocooler operation without any changes in the surface or size variations of the wear zones. This inferred that DLC-coated piston coolers can be easily operated at least for twice the operation hrs in comparison to TiN-coated piston coolers as given in Table 7. Also, the power drawn by the cryocooler was approximately half in comparison to the cryocooler with a coated piston.

The major outcome based on tribological evaluation and cooling performance assessment of hard-coated stirling cooler pistons is that DLC coating exhibited better wear resistance than TiN. The DLC-coated pistons cryocooler gave a satisfactory cooling performance for even more than 300 hrs of continuous cycle tests without any wear on the piston surface. DLC surface coating emerged as a more suitable coating on cryocooler pistons for the increased operational life of the Stirling cryocooler in field operation.

References

[1] S. Zhang and W. Zhu, Tin coating of tool steels: A review. J. Mater. Process. Technol., *39*(1–2) (1993) 165–177. https://doi.org/10.1016/0924-0136(93)90016-y.
[2] R.J. Rodrìguez, Tribological behaviour of hard coatings deposited by arc-evaporation PVD. Vacuum, *67*(3–4) (2002) 559–566. https://doi.org/10.1016/s0042-207x(02)00248-8.
[3] Deng Jianxin and Liu Aihua, Dry sliding wear behavior of PVD Tin, $Ti_{55}Al_{45}N$, and $TI_{35}AL_{65}N$ coatings at temperatures up to 600 °C. Int. J. Refract. Met. Hard Mater., *41* (2013) 241–249. https://doi.org/10.1016/j.ijrmhm.2013.04.008.
[4] S.Y. Lee and S.H. Kim, Metal nitride coatings by physical vapor deposition (PVD) for a wear resistant aluminum extrusion die. J. Nanosci. Nanotechnol., *14*(12) (2014) 8993–8998. https://doi.org/10.1166/jnn.2014.10058.
[5] A. Baptista, F. Silva, J. Porteiro, J. Míguez and G. Pinto, Sputtering physical vapour deposition (PVD) coatings: a critical review on process improvement and market trend demands. Coatings, *8*(11) (2018) 402. https://doi.org/10.3390/coatings8110402.
[6] A. Aytac, U. Ozturk and H. Sert, Effect of TiN coating on wear behaviour of mould punch. J. Balkan Tribol. Assoc., *19*(3) (2013) 433–447.
[7] J.A. Ghani, I.A. Choudhury and H.H. Masjuki, Wear mechanism of tin coated carbide and uncoated cermets tools at high cutting speed applications. J. Mater. Process. Technol., *153–154*(2004) (2004) 1067–1073. https://doi.org/10.1016/j.jmatprotec.2004.04.352.
[8] S. Koseki, K. Inoue, S. Morito, T. Ohba and H. Usuki, Comparison of tin-coated tools using CVD and PVD processes during continuous cutting of ni-based superalloys. Surf. Coat. Technol., *283* (2015) 353–363. https://doi.org/10.1016/j.surfcoat.2015.10.071.
[9] G. Xian, J. Xiong, H. Fan, F. Jiang, Z. Guo, H. Zhao, L. Xian, Z. Jing, J. Liao and Y. Liu, Investigations on microstructure, mechanical and tribological properties of tin coatings deposited on three different tool materials. Int. J. Refract. Met. Hard Mater., *102* (2022) 105700. https://doi.org/10.1016/j.ijrmhm.2021.105700.
[10] C.Y. Hsu, C.C. Chiang, T.S. Hsieh, H.C. Hsu, L. Tsai and C.H. Hou, Study of fiber bragg gratings with tin-coated for cryogenic temperature measurement. Opt. & Laser Technol., *136* (2021) 106768. https://doi.org/10.1016/j.optlastec.2020.106768.
[11] J. Liu, Y. Zhang and B. Liao, A review on preparation process and tribological performance of coatings for internal combustion

engine piston ring. Adv. Mech. Eng., *15*(5) (2023) 1–19. https://doi.org/10.1177/16878132231175752.
[12] S.V. Johnston and S.V. Hainsworth, Effect of DLC coatings on wear in automotive applications. Surf. Eng., *21*(1) (2005) 67–71. https://doi.org/10.1179/174329405x30039.
[13] A. Götze, S. Makowski, T. Kunze, M. Hübner, H. Zellbeck, V. Weihnacht, A. Leson, E. Beyer, J.O. Joswig, G. Seifert, G. Abrasonis, M. Posselt, J. Fassbender, W. Möller, S. Gemming and M. Krause, Tetrahedral amorphous carbon coatings for friction reduction of the valve train in internal combustion engines. Adv. Eng. Mater., *16*(10) (2014) 1226–1233. https://doi.org/10.1002/adem.201400188.
[14] M. Kano, Diamond-like carbon coating applied to automotive engine components. Tribol. Online, *9*(3) (2014) 135–142. https://doi.org/10.2474/trol.9.135.
[15] K. Bobzin, T. Brögelmann, C. Kalscheuer, M. Thiex, M. Ebner, T. Lohner and K. Stahl, A contribution to the thermal effects of DLC coatings on fluid friction in EHL contacts. Lubr. Sci., *30*(6) (2018) 285–299. https://doi.org/10.1002/ls.1421.
[16] A. Vanhulsel, F. Velasco, R. Jacobs, L. Eersels, D. Havermans, E.W. Roberts, I. Sherrington, M.J. Anderson and L. Gaillard, DLC solid lubricant coatings on ball bearings for space applications. Tribol. Int., *40*(7) (2007) 1186–1194. https://doi.org/10.1016/j.triboint.2006.12.005.
[17] M. Kalin and J. Vižintin, The tribological performance of DLC-coated gears lubricated with biodegradable oil in various pinion/gear material combinations. Wear, *259*(7–12) (2005) 1270–1280. https://doi.org/10.1016/j.wear.2005.02.028.
[18] S.V. Hainsworth and N.J. Uhure, Diamond Like Carbon Coatings for tribology: Production techniques, characterisation methods and applications. Int. Mater. Rev., *52*(3) (2007) 153–174. https://doi.org/10.1179/174328007x160272.
[19] B. Podgornik and J. Jerina, Surface topography effect on galling resistance of coated and uncoated tool steel. Surf. Coat. Technol., *206*(11–12) (2012) 2792–2800. https://doi.org/10.1016/j.surfcoat.2011.11.041.
[20] P. Hanzelka, T. Kralik, A. Maskova, V. Musilova and J. Vyskocil, Thermal radiative properties of a DLC coating. Cryogenics, *48*(9–10) (2008) 455–457. https://doi.org/10.1016/j.cryogenics.2008.03.021.
[21] G. Walker, R. Fauvel and G. Reader, Miniature refrigerators for cryogenic sensors and cold electronics. Cryogenics, *29*(8) (1989) 841–845. https://doi.org/10.1016/0011-2275(89)90158-6.
[22] J. Fereday, T. Bradshaw, M. Crook, A. Orlowska, R. Bhatia, M. Linder, O. Pin, A. Scommegna and S. Vey, Cryocooler modelling methodology. Cryogenics, *46*(2–3) (2006) 183–190. https://doi.org/10.1016/j.cryogenics.2005.10.003.
[23] M.M. Lele, S.L. Bapat and K.G. Narayankhedkar, performance evaluation of stirling cycle cryocoolers, Cryo Pragaue 2006, 21st International Cryogenic Engineering Conference (ICEC21), Czech Republic (2006)
[24] M.D. Atrey, S.L. Bapat and K.G. Narayankhedkar, Optimization of design parameters of stirling cycle machine. Cryogenics, *33*(10) (1993) 951–957. https://doi.org/10.1016/0011-2275(93)90223-b.
[25] S.J. Park, Y.J. Hong, H.B. Kim, D.Y. Koh, J.H. Kim, B.K. Yu and K.B. Lee, The effect of operating parameters in the stirling cryocooler. Cryogenics, *42*(6–7) (2002) 419–425. https://doi.org/10.1016/s0011-2275(02)00063-2.
[26] J.M. Pfotenhauer, R.Z. Wang and F.K. Miller, Regenerator design optimization: Results from regen 3.3. Cryogenics, *97*(2019) (2019) 77–84. https://doi.org/10.1016/j.cryogenics.2018.11.009.
[27] S. Chen, Q. Huang, M. Liang, H. Chen, L. Chen and Y. Hou, Numerical study on the heat transfer characteristics of oscillating flow in cryogenic regenerators. Cryogenics, *96* (2018) 99–107. https://doi.org/10.1016/j.cryogenics.2018.10.012.
[28] P.C.T. de Boer, Optimal performance of regenerative cryocoolers. Cryogenics, *51*(2) (2011) 105–113. https://doi.org/10.1016/j.cryogenics.2010.12.002.
[29] E.C. Landrum, T.J. Conrad, S.M. Ghiaasiaan and C.S. Kirkconnell, Hydrodynamic parameters of mesh fillers relevant to miniature regenerative cryocoolers. Cryogenics, *50*(6–7) (2010) 373–380. https://doi.org/10.1016/j.cryogenics.2010.01.017.
[30] M. Singh, M. Sadana, S. Sachdev and G. Pratap, Development of miniature stirling cryocooler technology for infrared focal plane array. Def. Sci. J., *63*(6) (2013) 571–580. https://doi.org/10.14429/dsj.63.5756.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 24 2024

Experimental analysis of GMAW welded AA7475 T7351 alloy joints

Hari Shanker ✉; Reeta Wattal

Check for updates

AIP Conf. Proc. 3111, 060008 (2024)
https://doi.org/10.1063/5.0221497

View Online
Export Citation

06 August 2024 04:46:25

Nanotechnology & Materials Science
Optics & Photonics
Impedance Analysis
Scanning Probe Microscopy
Sensors
Failure Analysis & Semiconductors



Unlock the Full Spectrum.
From DC to 8.5 GHz.
Your Application. Measured.
Find out more
Zurich Instruments

AIP Publishing

Experimental Analysis of GMAW Welded AA7475 T7351 Alloy Joints

Hari Shanker[1,a)], Reeta Wattal[2]

[1]*Department of Mechanical Engineering, Ajay Kumar Garg Engineering College Ghaziabad, Uttar Pradesh, 201009, India*
[2]*Department of Mechanical Engineering, Delhi Technological University, Delhi, 110042, India*

[a)] *Corresponding Author: chaursiyahs@akgec.ac.in*

Abstract: An alloyof AA7475 sheets and plates is currently being used for high-performance aircraft applicationsbut the main hurdle for such applications is its proper joining technique because AA7475 alloy is hard to weld. This investigation is done to characterize the microstructural and mechanical properties of a weld created by GMAW on AA7475 alloy in T7351 temper conditions. The tensile strength (TS) yields strength (YS), and percentage elongation was analysed and observed that theweld metal (WM) has a % elongation of 3% which is slightly less than 3.2% of heat affected zone (HAZ) and significantly less than 7.2% of the base metal (BM). Hardness tests were also performed and it was observed that the weld metal carried lower hardness values.For microstructural analysis, the micrograph and corresponding Energy Dispersive X-ray analysis were done at base metal, HAZ and weld metal. The compositions at the top, centre and root of the weld were discussed through EDX analysis. The fractography of WM ofthe fractured tensile samplewasalso studiedand reported.

Keywords: GMAW, Aerospace material, Aluminium Alloy, SEM, EDX.

06 August 2024 04:46:25

INTRODUCTION

The applications of lightweight material in weight-sensitiveapplications like aircraft, automobiles, and marine life are growing day by day. It is also observed that weight reduction is directly related to the reduction in fuel consumption and the economy of the products [1]. Initially, these applications are fulfilled by using wood, iron, and their derivatives which have high weight orless corrosion resistant as well as expensive. To replace these heavy and bulky parts and components, a material is required that can replace these materials with reasonable qualities. After rigorous research, it was found that aluminium and its alloy can replace these materials because it isinexpensive, lightweight has high resistance to corrosion, is easily recyclable,also heat treatable to a major level of strength [2]. As the research advances, highly tailored high-strength aluminium alloys are being created to fulfil the industry requirements [3]. As a result, different series of aluminium alloys were developed to fulfil the time-to-time requirements. In this regard,different series of aluminium alloys like 2000, 4000, 5000, 6000, and 7000 were developed to fulfil the demand of the industry. The addition of 2% of copper along withZn and Mg, is able tosignificantly increase the strength of AA7xxx series alloys [4]. The new AA7000 series alloy like AA7050 offers excellent corrosion resistance, improved fatigue, strength and toughness. The alloy AA7075 was introduced by changing its Si and iron content as well as by changing its quenching/ageing conditions, enhancing the fatigue strength and fracture toughness. In the AA7075 alloy, the total iron content is 0.90 % and by decreasing its iron content by 0.22 per cent a new aluminium alloy, AA7475 is formed with excellent corrosion resistance and high fatigue strength [5]. The AA7475-T7351 has greater fracture toughness but comparatively lower yield strength and significantly improved ductility than the AA7075-T651 [6]. In the low-stress region, the alloy AA7475-T7351 outperforms in comparison to other 7000 series aluminium alloys. Due to its superior properties nowadays, alloy AA7475 is frequentlyused as a sheet and plate in high-performance aerospace applications and for components working in a hostile environment. With these facts and properties, the heat-treated AA7475 alloy will surely and safely be able to reduce the weight of the spacecraft structure and simultaneously improve fuel economy [7].

It is observed that as material improves, to form a structure,improved joining methods are required. Welding is generally used as a connecting method to form a structure of aluminium, and GMAW isa generally usedmethod for joining Al-alloy parts and components in automotive, marine, and aerospace applications [8].The GMAW process shows high welding efficiency, and less time consumption, has a simple setup and is ultimately cheap in the fabrication process [9,10]. Despite many advantages, the base metal's mechanical and microstructural properties were changed during welding due tothe melting of base metal with a highly heated

AICTE & DSIR Sponsored International Conference on Smart and Innovative Development in Science, Engineering & Technology 2023
AIP Conf. Proc. 3111, 060008-1–060008-12; https://doi.org/10.1063/5.0221497
Published under an exclusive license by AIP Publishing. 978-0-7354-5008-0/$30.00

arcand if not controlled, provide poor weld joint properties. To fulfil the requirements of high strength and high corrosion-resistant material in different applications the joining of 5000 and 6000 series aluminium alloys was necessitated [11]. An AA6061 is welded by the GMAW process and concluded that the fracture takes place at the WM [12]. Defects have a significant effect on weld quality as well as joint efficiency. In this regard, different welding techniques have been tried to form an efficient weld with improved mechanical properties for different grades of Al-alloy and they are TIG, MIG, LBW, FSW, and GMAW. GMAW is a commonly used method for joining Al-alloy parts and components in the UK and also in Europe [13].Likedifficult-to-weld,all AA7000 series alloys as an age-hard-enable alloy,the aluminium alloy 7475 is also found very difficult to weld, by common fusion welding processes or by FSW [14].The GMAW process is cheaper and more flexible for joining thick sections compared to other welding processes [15]. The deeper penetration and high deposition rate were ensured by the GMAW process due to its continuous wire-feeding system and as a result, it produces high-strength weldments [16].

So, due to its distinct benefits over other welding processes,the GMAW welding process is considered the optimum solution to weld a such hard-to-weld, age-hard-enable Al-alloy 7475. In this investigation, the alloy of AA7475 in T7351 temper was welded by using the GMAW process, and the mechanical and microstructural properties were analysed tocheck the potential of the GMAW process for hard-to-weld, age-hard-enable Alalloy.

MATERIALS AND METHODS

Base Metal and Filler Material

An aluminium 7475 in T7351 temper condition sheet of thickness 3mm is taken as the base material. Coupons were cut and put in but joint configuration and are spaced at 1 mm.By using the protocol of ASTM-E-1251-17, the base material and filler compositions are reported in Table 1. Table 2 represents the mechanical properties of base Al-alloy. As per the ASTM E-8-16a Protocol,the mechanical propertiesof the AA7475-T7351 plate as BMare shown in Table 3.Table 4, represents the welding parameters that are used for the best weld bead after a set of pre-experiments. The oxide layers were removed before the weld with the help an emery paperand acetone.The argon as a shielding gas was used with a purity of 99.9% and fed with a 13.8 l/min flow rate, 3 seconds before the arc was struck.

Table1. Chemical Composition: Aluminium Alloy and filler metal by using, ASTM-E-1251-17 protocol:

Element	Results %	ER4043
Si	0.0507	5.4
Fe	0.0291	0.055
Cu	1.51	0.32
Mn	0.0072	0.76
Mg	2.23	0.047
Ni	0.0042	0.12
Zn	5.86	0.018
Ti	0.0188	-
Cd	<0.00010	<0.010
Co	<0.0010	<0.012
Pb	0.0112	-
Sn	0.0022	<0,004
Cr	0.211	-

06 August 2024 04:46:25

Table2. Mechanical properties of the Base Al-alloy by using Protocol: - ASTM E-8-16a:

YS (MPa)	UltimateT/S (MPa)	Fatigue Strength (MPa)	Hardness (HV)	Elongation (%)
370	435	160	170	7.5

Table 3. Details of Samples as per the Protocol, ASTM E-8-16a.

Specimen	Size (mm)	Area ($mm^{2)}$	GL (mm)	UT Load (KN)	Tensile Strength (N/mm^2)	0.2% PS Load (KN)	0.2%PS (N/mm^2)	Final GL (mm)	% E/L
1.	12.35 X 3.00	37.05	50.0	16.124	435	14.116	381	53.75	7.5

Table 4. Welding parameters optimised for best weld bead:

S. No.	Current (Amp)	Voltage (Volt)	WS(mm/s)	WFR (m/min)
1.	95	19.1	3.08	4.5
2.	80	18.9	4.06	3.9

EXPERIMENTAL SET-UP AND PROCEDURE

GMAW set-up and process-

A semi-automatic GMAW welding setup carrying a TPS 320i GMAW welding power source which has an inbuilt wire feeder, a welding torch, an air cleaner for an air gun, which is operated by an air compressor, a welding table equipped with anadjustable demonstration wall is shown in Figure 1. A spool of UNI-ER 4043 ($AlSi_5$) welding wire of diameter 1 mm is used as filler material. A number of pre-experiments were performed to optimize the welding parameters for the best weld bead.



FIGURE 1. Schematic of the Semi-automatic complete GMAW welding setup

Microstructural Observations:

A microstructural study is performed by extracting a specimen from a weld cross-section as shown in Figure 2. The specimen was etched in Keller solutionfor 35 seconds. BMI-101 A, an optical microscope was used to studythe microstructure of the joint, andby using aLEO EVO-40XVP, an SEM,the micrographs were taken and analysed. The fractured surface wasanalysedwith an SEM equipped with EDX to evaluate the compositions of a weld after performing tensile tests. BM, and WM, as well as HAZ, are analysed from SEM micrographs and corresponding phase composition with EDX.

06 August 2024 04:46:25

2.3 Mechanical Characteristics of the joint

2.3.1 Mechanical properties of the joints:

A non-destructive test was done and was followed by visual observation to identify the flaw left with the weld surface. A liquid penetrant test (LPU) was performed and itwas observed that there were no significant welding flaws available on the weld surface. A Vickers Hardness Tester (VH-1000B), was used to evaluate the microhardness of the transverse portion of the welded specimen by using ASTM E-384:11, ASME Sec-IX-2019 protocol. The tensile test samplesaremade as per the ASTME E-16a protocol. Thetensile strength, yield strength as well and % elongation was assessed on tensile test specimens, also the corresponding tensile samples andweld plateare shown in Figure 2(a, b).

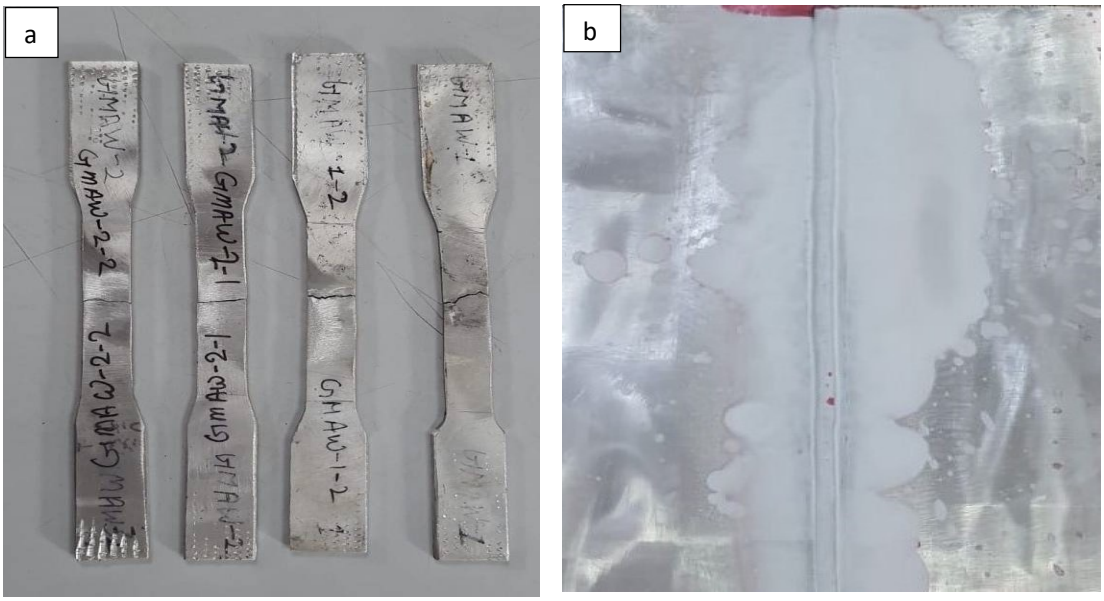


FIGURE 2. Shows the (a) welded samples for tensile testand (b) welded plate.

RESULT AND DISCUSSION

Weld Defect-

After visual inspection and liquid penetrant test, it is observed that there is no measure defect or other imperfections on the weld surface. From the macro-structure of the weld at 10X, some porosities are visible at the weld surface but they are few in number and have negligible effects on weld quality. There are no blow holes, hot cracks, orother harmful defects on the weld as per the test results as shown in Table 5.

Weld Structure Test:

Table 5. WM Test as per the Protocol of ASME S-9, ASTM E-381, ASM V-IX.

Parameters:			
Points of WeldNo.	**Height of Weld**	**Length of Weld**	**Cracks/Blow hole/Harmful Defect**
1.	3.251 mm	6.258mm	Not Observed at theWM

Macrostructures of weld sample

Figure 3, shows that there is no connection between the fusion area and the HAZ at even low magnification. There is a little difference in compositions between the weld metal and base metal except a minor concentration of Si elements on the weld metal, that may be due to the filler material.

06 August 2024 04:46:25



FIGURE 3. Macrostructures of welded specimen at a Magnification of - 10X**.**

Tensile properties of BM, WM, and HAZ

To examine the weld performance, a tensile test was performed at the BM, HAZ, and WM. The TS, and YS, as well as % elongation, was measured and the values are presented in Figure 4.Among all, the best tensile curvesarechosen and compared. In terms of % elongation, the BM shows 7.5% in comparison to 3% of WM and 3.2% of HAZ. It observed that the weld metal does not experience plastic deformation as compared to BM and HAZ and as a result, the WM shows lowest elongation. Due to this, there is a loss of strength, an increase in flexibility and significant softening. GMAW welded AA6061 joint [17] confirms that the weld metal showed lower tensile strength than BM and HAZ also showed lower plastic deformation.The tensile strength of BM is 435 MPa, much greater than 220 MPa of HAZ and 218 MPa of WM. Similar patterns are observed in the case of YS. All the specimen fails at the weld metal. The TS and YSvalues of HAZ are less than that of the BM but more than that of the WM, this may be due to the coarsening that takes place inside the HAZ. HAZ works as a heat sink probably due to the faster cooling rate that takes place at the weld metal and HAZ appears to be coarser than weld metal [18]. It was found that for the stress value 0 to 435 MPa of the base metal the load-displacement curve is linear and during that, the specimenexhibitsa pattern of elastic deformations. After that,the sample shows a homogeneous plastic deformation with an increase in stresses till the fracture takes place. Figure 4 (a-c) shows the load-displacement curve of the base metal, HAZ and weld metal. HAZ load-displacement curve is drastically different as compared to BM as shown in Figure 4b. As the load increase, the material starts yielding rapidly and get distorted followed by eventual breaking due to plastic slip deformation experienced by the material [19]. The WMengineering displacement curve is shown in Figure 4c. In this, it is observed that the WM shows a significantly lower value of percentage elongation than the HAZ and base metal, this is due to plastic deformation that takes place till the breaking point. Table 6 shows the average values of TS and % elongation for the WM and heat-affected zone.WM has a % elongation of 3% which is slightly less than HAZ and significantly less than BM.So, it is concluded that in terms of % elongation, TS, and YS, the WM is the weakest link of the weldments formed and all the samples failed at the weld metal.

Table 6. TS and % Elongation Values of the WM and HAZby following the protocol ofASTM E-8-16-a.

Samples	Size (mm)	Area (mm^2)	GL (mm)	UT Load (KN)	TS (N/mm^2)	Final GL (mm)	% E/L	Fracture Locations
1.	12.40 X 2.98	36.95	50.0	8.052	218	51.57	3.0	WM
2.	12.40 X 2.95	36.58	50.0	8.052	220	51.47	3.2	HAZ

06 August 2024 04:46:25

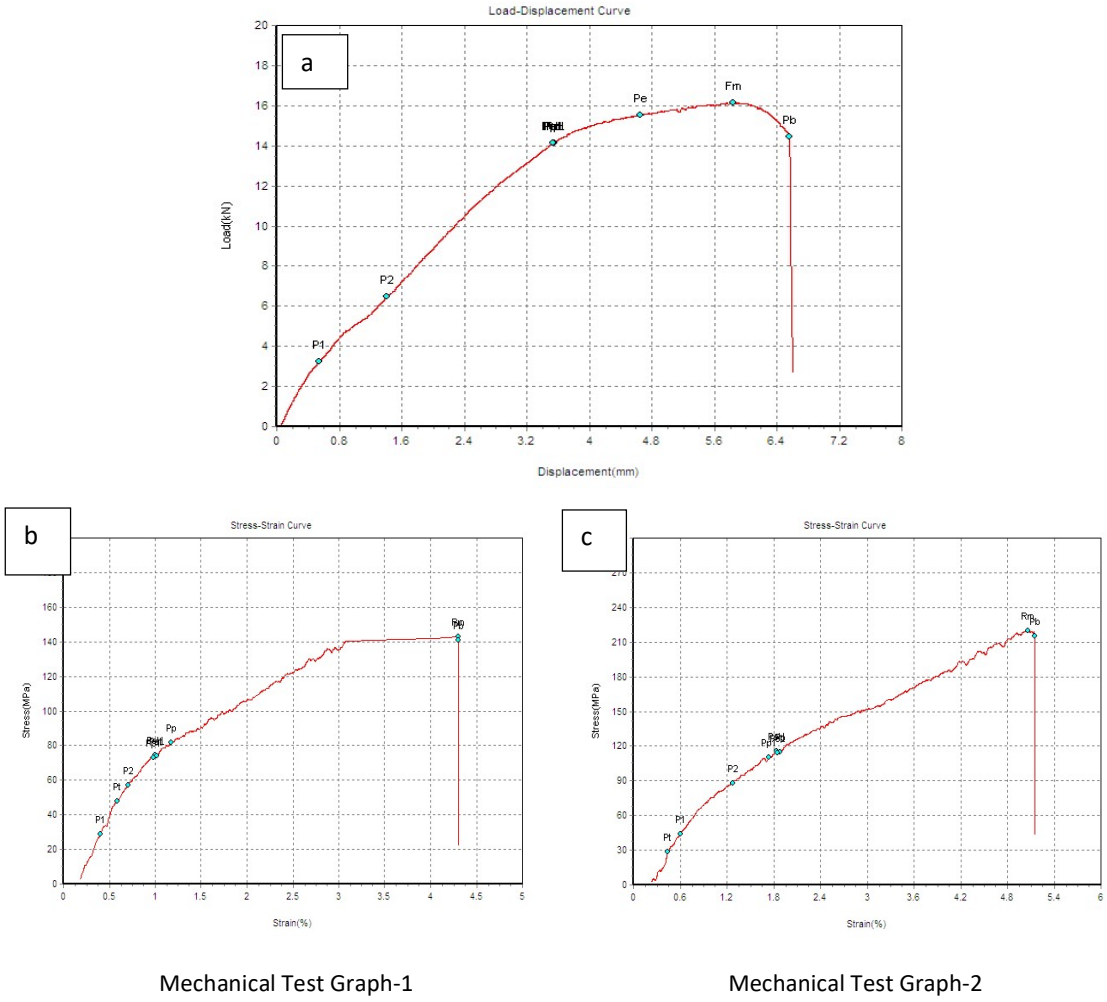


FIGURE 4. A Load displacement curve of (a) BM (b) HAZ (c) WM forAA7475 T-7351 alloysamples.

Microhardness and Microstructure of Base Metal, HAZ, Weld Metal

The weld characteristics were studied for theAA7475-T7351 alloy welded joint, a harness test was conducted by using the protocol of ASTM E-384-11, ASME Sec-IX-2019 and a micrograph was taken from the BM, HAZ, and WM of the specimen. Table 7, illustrates the weld metal and HAZ microhardness test results. The WM shows the lowest hardness which is 57.64 percent of the BM. The recorded hardness value of 170 HV was observed at BM which is significantly larger than that of the 98 HV of WM. The WM is primarily made with the filler metal and shows a heterogeneous structure this may be the cause of the lower hardness value as compared to BM. Al5.5Zn2.5Mg2.2Cu welded joints, have observed similar outcomes where the weld metalis observed as the weak link and carrying hardness only 56 % of BM. This may be due to movements of alloy constituents towards the grain boundaries and softening of the WM at the time of solidification [20]. This may be due to the softer filler wire material than BM.Columnar grain is the outcome of a joint made by ER4043 wire whereas ER5356 provides fine equiaxed grains. Fine equiaxed grains result in improved mechanical properties [21].Stainless steel and copper are successfully welded by the GMAW process and microhardness results reveal that the reduction in microhardness was observed at HAZ and WZ as compared to the base metal on both sides [22].The hardness value of HAZ is 136 HV which is 80% of the BM.This is due to the fact that, the HAZ is made with the help of semi-molten base metal.

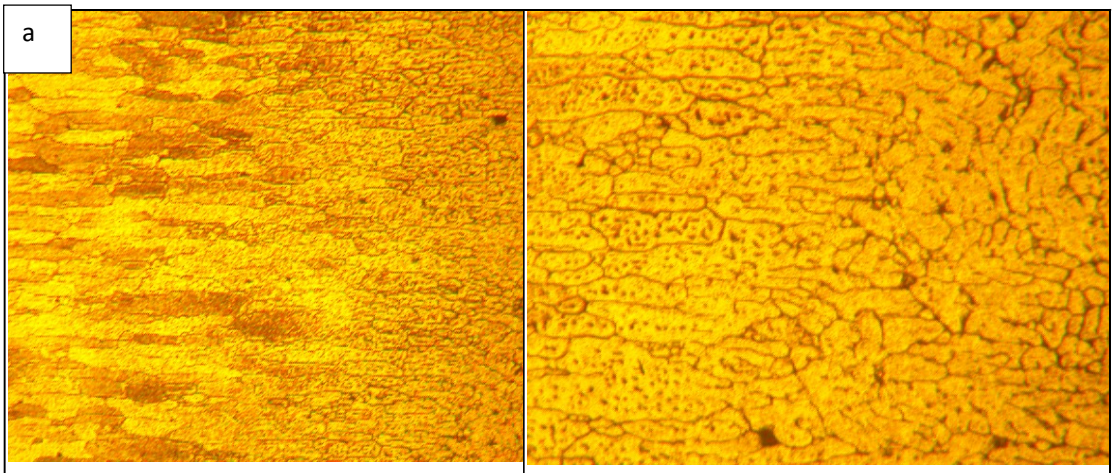
A typical metallographic image was takenfrom the BM, HAZ as well and WM by using ASME Sec-9, ASTM E-381, and ASM V-9 Protocol and aredisplayed in Figure 5(a-c).The base metal shows elongated grains at 400X. The dendritic structure was reported due to melting and solidification in the weld metal during GMAW welding which shows elongated grains [23].Figure 5a depicts the microstructure of the HAZ at 100X and 200X. The General Microstructure shows ununiformly dendritic structure observed throughout the matrix.At the weld

06 August 2024 04:46:25

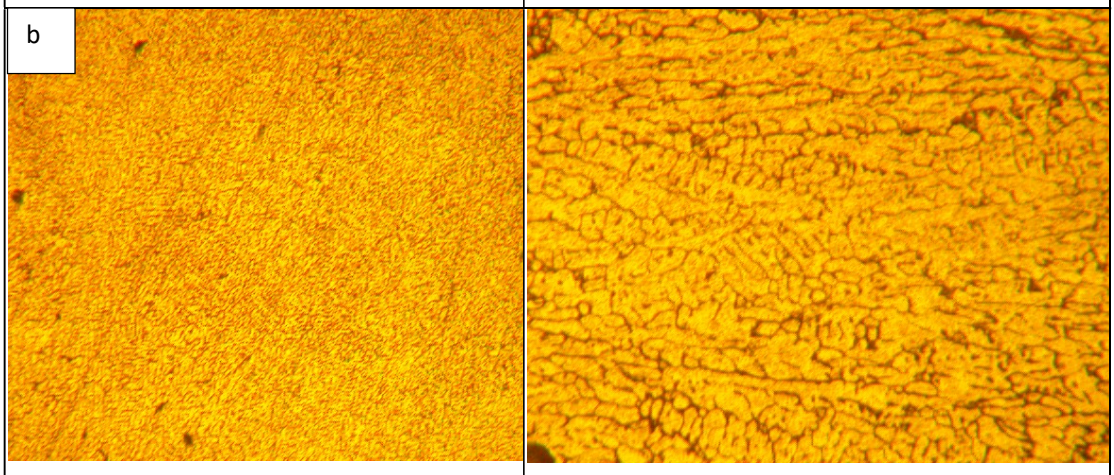
interface, a dendritic microstructure was observed that is due to the high cooling rate and less heat input. [24,25].Due to thermal cycling, the columnar crystallization takes place at the molten pooland shows typical eutectic crystallization behaviour. Figure 5 shows the weld metals metallographic image at a magnification of 100X and 200X. The structure shows a fine and dendritic grain and is observed as full penetration welding in the BM.The metallograph in Figure 5c at 400X shows the HAZ and WMmicrostructures which havea non-uniform and uniform dendritic grain structure respectively. A faster cooling rate refined the grain size because of no time of grain growth and shorter grains are formed [26].The HAZ shows a larger and coarser grain structure compared to weld metal which may be due to the high cooling rate [17].

Table 7. Hardness Test of WM and HAZ by usingASTM E-384/11,ASME SEC-9-2019 protocol

Parameter	**Observed Value(HV)**	**Result(Average)**
Hardness test (WM)	99,97,97	98HV
Hardness test (HAZ)	140,136,131	136HV

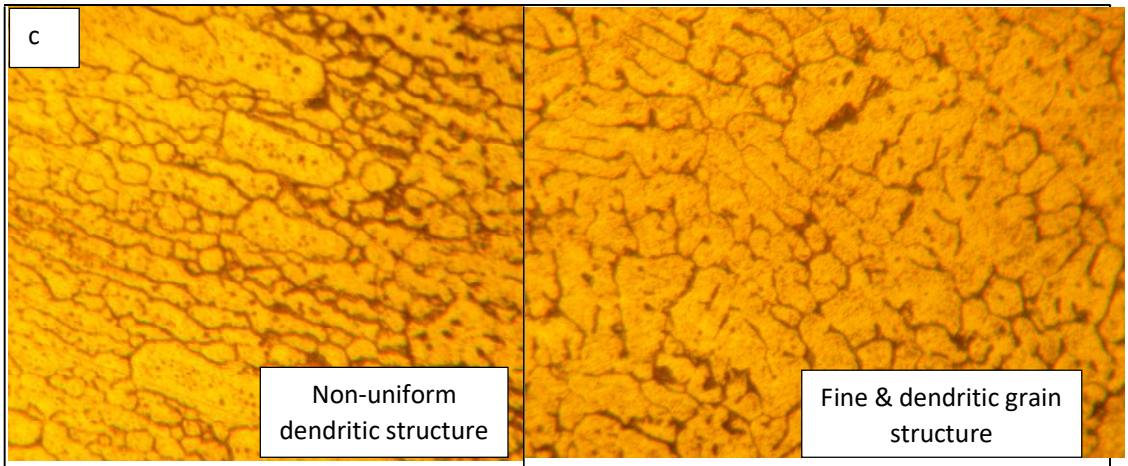


HAZ Microstructure	HAZ Microstructure
Mag- 100X	Mag- 200X



WeldMicrostructure	WeldMicrostructure
Mag- 100X	Mag- 200X

06 August 2024 04:46:25

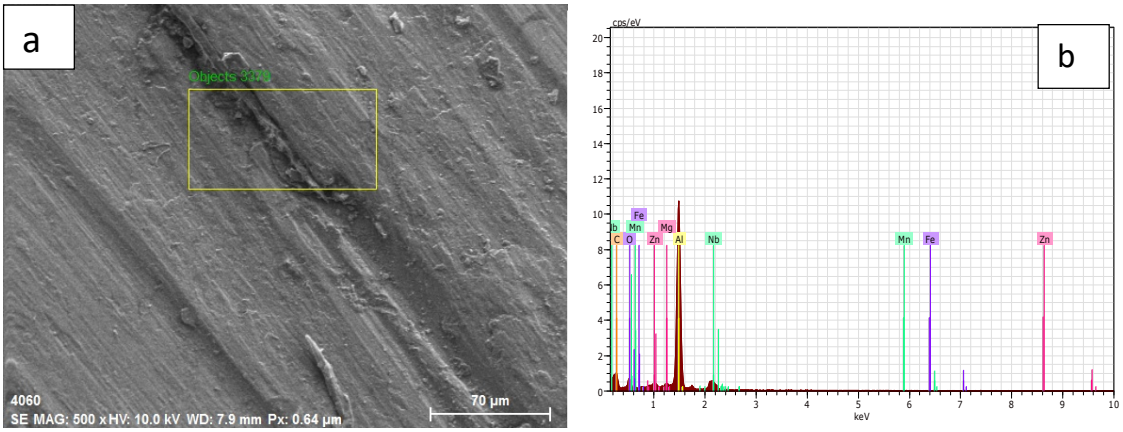


HAZ Microstructure	WeldMicrostructure
Mag- 400X	Mag- 400X

FIGURE 5. Metallography of GMAW Weld, (a) BM (b) HAZ (c) HAZ and WM, at different magnifications.

Microstructural composition analysis of weld

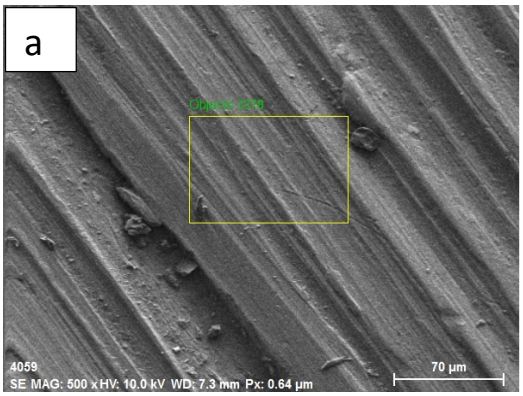
The above hardness values and the tensile test resultshow that in comparison to the HAZ and BM, the WM is the weakest link. So, for the detailed analysis of the weld EDX analysis is done with a corresponding SEM micrograph taken from the top middle and root of the weld fractured surface are shown in Figure 6 (a-c). EDX result in Figure 6a shows that the constituent Mg, Fe, Cu, Zn and other elements were migrated from the BM towards the WM during the welding process. The results of EDX from the top portion of the weld reveal 47.47% Al, 25.99% Cu,2.16% Zn, and 8.04% O elements, indicating that the solution is an Al-Cu-based solid solution. In Figure 6-b which is from the weld centre of the weld the EDX result shows an increment in Al elements up to 65.43% and a decrease in the Cu and O elements by 15.02% and 4.23 % respectively. This indicates that the centre of the weld is not fully made up of an Al-based solid solution. In this, there is a diffusion of some other elements like Mg, Fe and Zn from base metal to WZ enhanced. The composition from the root of the weld shows pure Al-based solid solution with Al-80.97%, O element 4.51 % and an increase in the Zn and Mg content was observed. The diffusion of Cu content was not found in this region and is shown in Figure 6c. Similar outcomes are observed and explained in the solidification theory that at the solid-liquid interface an accumulation layer was formed by varying amounts in the welding pool [27]. Based on particular research, the particle phase of the Al/Cu reaction was primarily made up of Cu_2Mg. As a result, it can be explained that at the top and centre of the fusion zone, there may be some brittle inter-metallic compounds dispersed that are the cause of the fracture[28].Similar outcomes werefound for all the analysed welded samples.

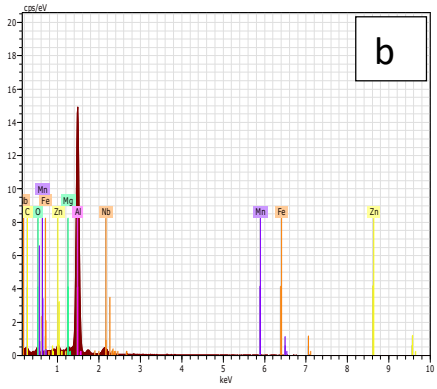


06 August 2024 04:46:25

Spectrum of the: Objects 3379					c
E1ANSeriesunn.	C norm.	C Atom.	C Error (1 Sigma)		
	[wt.%]	[wt.%]	[at.%]	[wt.%]	
Al 13 K-series	19.71	47.47	37.88	0.97	
C 6 K-series	10.79	25.99	46.60	3.05	
Nb 41 L-series	6.30	15.17	3.52	0.43	
O 8 K-series	3.34	8.04	10.83	1.10	
Zn 30 L-series	0.90	2.16	0.71	0.14	
Mn 25 K-series	0.31	0.75	0.29	0.20	
Fe 26 K-series	0.16	0.39	0.15	0.17	
Mg 12 K-series	0.01	0.02	0.02	0.03	
Total:	41.52	100.00	100.00		

FIGURE 6. An EDX results with SEM micrographs of weld samples from Top (a) SEM Image (b) EDX Graph (c) Elemental Analysis





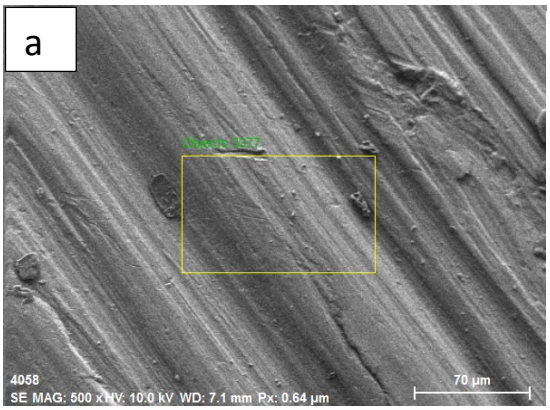
Spectrum: Objects 3378					c
El AN Seriesunn.	C norm.	C Atom.	C Error (1 Sigma)		
	[wt.%]	[wt.%]	[at.%]	[wt.%]	
Al 13 K-series	38.08	65.43	58.21	1.77	
C 6 K-series	8.74	15.02	30.01	2.36	
Nb 41 L-series	5.55	9.54	2.46	0.33	
O 8 K-series	2.46	4.23	6.35	0.74	
Zn 30 L-series	1.43	2.45	0.90	0.15	
Fe 26 K-series	0.81	1.39	0.60	0.26	
Mg 12 K-series	0.66	1.13	1.11	0.08	
Mn 25 K-series	0.48	0.82	0.36	0.17	
Total:	58.21	100.00	100.00		

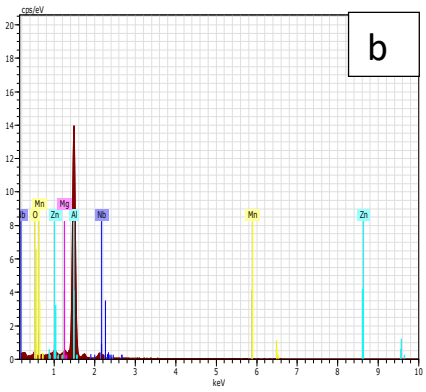
FIGURE 7. An EDX results with SEM micrographs of weld samples from Centre (a) SEM Image (b) EDX Graph (c) Elemental Analysis

Fracture Analysis of Weld

The fracture behaviour of the weld surface after the tensile test was analysed using SEM. For this, the SEM fractographywas taken from the top, centre, and root (bottom) of the weld after failure under tensile loading and presented in Figures7(a-c).It is observed that there were no visible defects over the weld surface although the fracture took place at WM. Fractography shown in Figure 7a taken from the weld top portion reveals a fracture with a large number of dimples that are equiaxed and uniformly distributed. Also, the fractography from the middle of the weld shows, that the dimples are nearly identical with varying numbers of fine dimples on the fractured surface of the weld shown in Figure 7b. Through the fractography from the top and centre of the fractured weld surface, it can be concluded that the major portion of failure is ductile in nature and that is due to the predomination of the dimples on the fractured surface. The flat surfaces represent the brittle fracture and the ductile fractures are represented by dimples. Also, some flat surfaces and dimple mixtures shows both ductile

06 August 2024 04:46:25





c

Spectrum: Objects 3377

El	AN	Series	unn. C [wt.%]	norm. C [wt.%]	Atom. C [at.%]	Error (1 Sigma) [wt.%]
Al	13	K-series	41.57	80.97	85.62	1.96
Nb	41	L-series	5.05	9.84	3.02	0.37
O	8	K-series	2.32	4.51	8.05	0.83
Zn	30	L-series	1.04	2.52	0.88	0.14
Mg	12	K-series	0.82	1.60	1.88	0.10
Mn	25	K-series	0.54	1.06	0.55	0.24
		Total:	51.34	100.00	100.00	

FIGURE 8. An EDX results with SEM micrographs of weld samples from Root. (a) SEM Image (b) EDX Graph (c) Elemental Analysis

and brittle fracture [29]. This conclusion is also supported by [30], in his findings he explained that the samples withheat-treated by solution have a sufficientquantityof fine dimples as compared to samples without heat-treated. Figure 7c shows the fractography taken from the rootof the fractured weld surface. It is observed that when the material goes under tensile test, voids emerge in the ductile material prior to necking and as a result, the cleavage failure pattern was the outcome. The coarse and elongated grain was observed at the root of the weld due to the early start of necking which resulted in much more evidence of void development. This outcome is supported by the pulsed wave GMAW joint created in PWHT samples and observed prolonged and coarse dimples [31]. At the weld root section, the metal consolation was generally inadequate this may be due to the poor weld quality and lack of penetration are the cause of brittle fracture that takes place in the root potion [32].

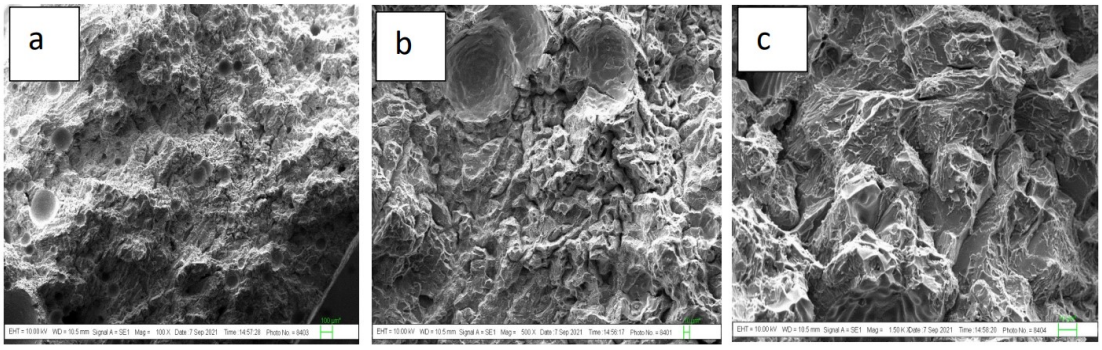


FIGURE 9. The SEM fractography oftensile tested weld sample; from (a) top; (b) centre and (c) root.

At the centre of the weld fracture surface the dimple appears wider and deeper in comparison to the sample from the top of the weld, as a result,it was concluded that the top portion of the weld is more ductile. The toughness fracturing mode is observed at the middle and root of the weld [33].Small and deep dimples represent the high ductility of the fractured tensile samples and the distance between the grain is the measure of ductility of the weld [34]. Cleavage failure patterns were discovered at the root of the weld.

06 August 2024 04:46:25

CONCLUSIONS

The main conclusions are summarized as follows:

1. WM has a % elongation of 3% which is slightly less than 3.2% of HAZ and significantly less than 7.2% of the BM. It is found that in terms of % elongation, TS, and YS, the weld metal is the weakest link of the weldments formed and all the samples failed in the WM.
2. The BM is the hardest portion with a harness of 170 HV, much greater than the HAZ and WM, and the lowest harness 98 HV, was recorded at the WM, which is only 57.64 % of the BM.
3. A metallographic image taken from BM shows that it is made up of elongated grains whereas an ununiform dendritic structure was observed in the HAZ throughout the matrix. The general microstructure of WMshows full penetration welding in the base structure. The structure consists of fine and uniform dendritic grain structure.
4. From EDX analysis it was found that during the welding process, the elements Mg, Zn, and Cu migrate from the BM towards the WM. The results from the top of the weld show a pure Al-Cu-based solid solution, at the weld centre, indicating that this location had not yet been made up of an Al-based solid solution, however, the diffusion of other elements like Zn, Mg, and Fe from BM to WM had enhanced. The EDX results from the root of the weld show a pure Al-based solid solution.
5. Fractography from the top of the weld depicts a fracture with a large number of equiaxed dimples which are uniformly distributed. The dimples here on the centre of the weld's fracture surface are wider and deeper than the dimples on the fracture surface of the top of the weld, showing that the top region of the weld is more ductile than the centre and root of the weld. Cleavage failure patterns were observed at the root of the weld.

ACKNOWLEDGEMENTS

The authors would like to acknowledge that the research work is completely his own work and purely for the requirements of the PhD degree from Delhi Technological University Delhi. Also, the author would like to confirm that there are not any financial agencies available to support the research work.

REFERENCES

1. F. C. Campbell Manufacturing technology for aerospace structural materials. Butterworth-Heinemann Publication, An Imprint of Elsevier Publications,(2006), New York, NY, USA.
2. J. Zehnder, in Aluminium: Technology, application, and environment, a profile of a modern metal (ed) Dietrich G Altenpohl (Pennsylvania: TMS Warrendale),(1996) p. 319.
3. B.B. Verma, TCT report (Sheffield: Sheffield Hallam University)1997.
4. S. H.Avner, Introduction to physical metallurgy (New Delhi: McGraw-Hill Education (India) Pvt Limited, (1997).
5. M. P.Groover, Fundamental of modern manufacturing, materials, processes and systems. 3rd Edition. John Wiley and Sons; 2007.
6. D. Tadjiev, H. T. Yang, S. T.Kim, Fatigue life prediction under random loading conditions in 7475–T7351 aluminum alloy using the RMS model. Int J Damage Mech (2006); 15:89–102.
7. M. F. Gittos,W.Xu, Material and structural behaviour of MIG butt weld in 6005- T6 aluminium alloy extrusion under quasi-static and impact loading. TWI report No. 14054/1/04; (2004).
8. G. Peng, K. Chen, L. Jia, P. Dong, S. Chen, Effect of heat treatment on strength, exfoliation corrosion and electrochemical behavior of 7085 aluminum alloy. Mater Des (2012); 35:93–8.
9. M. Ketabchi, S.E. Mirsalehi,S.Aliakbari, J. Manufact. Process. 35 (2018) 226– 232.
10. L. Huang, et al. Int. J. Adv. Manuf. Technol. 93 (2017) 1809–1817
11. J. Failu, K. N. Pandey, R. P. Verma, Anal. Prevent. 17 (2017) 248–254.
12. A. K.Lakshminarayanan, K. Elangovan,V. Balasubramanian, Int. J. Adv. Manuf. Technol. 40 (2009) 286–296.
13. C. Navarro, J. Vazquez, J. Dominguez, Experimental results in fretting fatigue with shot and laser peened Al 7075-T651 specimens. Int J Fatigue (2012); 40:143–53.
14. B. Hu, I. M. Richardson, Microstructure and mechanical properties of aa7075(t6) hybrid laser/GMA welds. Mater Sci Eng A (2007); 459:94–100.
15. D. I. Adebiyi,O.O.Gbadeyan, T. E. Abioye, "Analysis of the mechanical properties and penetration depth of gas metal arc welding on AISI 304 stainless steel", International Journal of Microstructure and Materials Properties, (2019), Vol. 14 No. 1, pp. 47-59.
16. A. Y. Ahmed, Al-Quensi. "Fusion welding techniques", International Journal of Engineering Research and Applications, (2016), Vol. 6 No. 3, pp. 78-83.
17. M. A. Bakar, R. Ahmad, "Effect of post-weld heat treatment on the mechanical and microstructure properties of AA6061 joints welded by the gas metal arc welding cold metal transfer method, Materials and Design 32 (2011) 5120–5126.

06 August 2024 04:46:25

18. I. S. Omotehinse, T. E. Abioye, Effects of post-weld heat treatments on the microstructure, mechanical and corrosion properties of gas metal arc welded 304 stainless steel, World Journal of Engineering 17/1 (2020) 87–96 © Emerald Publishing Limited [ISSN 1708-5284], (2019), [DOI 10.1108/WJE-11-2019-0323].
19. A. Elrefaey, Effectiveness of cold metal transfer process for welding 7075 aluminium alloys Science and Technology of Welding and Joining 20 280-85.
20. M. G. Reddy, V. Balasubramanian, V. Ravisankar V, Effect of pulsed current and post-weldageing treatment on tensile properties of argon arc welded high strength aluminium alloy. Mater Sci Eng A (2007); 459:19–34.
21. K. N. Pandey,R. P. Verma, and Y. Sharma, Effect of ER4043 and ER5356 filler wire on mechanical properties and microstructure of dissimilar aluminium alloys, 5083-O and 6061-T6 joint, welded by the metal inert gas welding J Engineering Manufacture (2014) 1–8 journalsPermissions.nav DOI: 10.1177/0954405414535771.
22. M. Zhang, Y. Zhang, et al. Microstructure and Mechanical Properties of the Joint Fabricated Between Stainless Steel and Copper Using Gas Metal Arc Welding Trans Indian Inst Met (2021) 74(4):969–978, https://doi.org/10.1007/s12666-021-02208-7.
23. A. Verma, B. Kotteswaran, et al. Microstructure and Mechanical Properties of AA7005 Alloy Joint by Fusion and Solid-State Welding Processes, Trans Indian Inst Met (2020) https://doi.org/10.1007/s12666-020-01917-9.
24. H. Queiroz, J. Duerst, N. Switzner, Mater Sci Eng A 709 (2018) 55.
25. A. K. Lakshminarayanan, S. Ramachandran, Trans Nonferr Met Soc China 30 (2020) 727.
26. C. Kose and C. Topal, "Effect of post-weld heat treatment and heat input on the microstructure and mechanical properties of plasma arc welded AISI 410S ferritic stainless steel", Materials Research Express, (2019), Vol. 6, p. 66517.
27. G. Straffellini, L. Bertini, V. Fontanari, Influence of post-weld treatments on the fatigue behaviour of al-alloy welded joints. Int J Fatigue (1998); 20:749–55.
28. H. Shanker and R. Wattal," A comprehensive survey on the cold metal transfer process in welding similar and dissimilar materials and for cladding, Proc IMechE Part C: J Mechanical Engineering Science, 2023, Vol. 237(10) 2360–2391, https://doi.org/10.1177/09544062221139947.
29. A. K. Gogia, R. Bobbili, V. Madhu, "Tensile behaviour of aluminium 7017 alloys at various temperatures and strain rates", Journal of Materials Research and Technology, (2019),Vol. 5 No. 2, pp. 190-197.
30. M. A. Bakar, R. Ahmad, Effect of post-weld heat treatment on the mechanical and microstructure properties of AA6061 joints welded by the gas metal arc welding cold metal transfer method. Materials and Design 32 (2011) 5120–5126. doi: 10.1016/j.matdes. (2011).06.007.
31. H. P. Burger, L. Ivanchev, R. Akhter, Effect of pre/post t6 heat treatment on the mechanical properties of laser welded ssm cast a356 aluminium alloy. Mat Sci Eng A (2007); 447:192–6.
32. F. Karimzadeh, M. A. Golozar, M. Heidarbeigy, Z. Nikseresht, Effect of heat treatment on microstructure and corrosion behaviour of al6061 alloy weldment. Mater Des (2010); 31:2643–8.
33. H. Shanker and R. Wattal, 'Mechanical and microstructure characteristics of 7475-T7351 aluminium alloy joints welded by robot cold metal transfer. ***Proc IMechE*** Part B: J Engineering Manufacture (2023) *1–13,* https://doi.org/10.1177/09544054231178956.
34. S. B. Jamaludin, M. H. Abdullah et al. "Effect of solution treatment temperature on tensile strength of Al-Mg-Si alloy", Materials Science Forum, (2015), Vol. 819, pp. 39-44.

06 August 2024 04:46:25

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575153

"Explainable AI" Disease Detection with Reasoning

1st Manoj Kumar
mkprajapat427@gmail.com
Department of Computer Engineering
Delhi Technological University
Bawana Road, Delhi

2nd Kanishk Ramrakhiyani
kanishkramrakhiyani@gmail.com
Department of Computer Engineering
Delhi Technological University
Bawana Road, Delhi

3rd Harsh Garg
garg28harsh@gmail.com
Department of Computer Engineering
Delhi Technological University
Bawana Road, Delhi

***Abstract*—Explainable AI (XAI) has gained significant attention globally due to its capability to provide transparency and understanding of complex machine learning models.We explore the potential benefits of XAI and integrate them into the domain of disease detection. Our objective is to leverage the power of large language models coupled with advanced explainability algorithms to enhance the accuracy and interpretability of disease predictions based on input symptoms. By doing so, we aim to not only achieve precise disease identification but also offer clear and comprehensible explanations for the predictions made. This approach holds the promise of bridging the gap between the intricate nature of AI-driven disease detection systems and the need for transparency and trustworthiness in medical diagnoses.**

***Index Terms*—SHAP, explainable AI,Large language models, Neural networks,shapley scores,Prompt development,Acceptance Score, Agreement Score, Weightage**

I. Introduction

The integration of Artificial Intelligence (AI) in healthcare has transformed disease detection and diagnosis. However, the adoption of complex machine learning models presents challenges in terms of transparency and explainability. Decision-makers, including medical professionals and patients, demand clarity and comprehensibility in AI-driven medical predictions to trust these systems.

The importance of "Explainable AI" (XAI) is highlighted by recent regulations, such as the proposed EU legislation mandating a "right to an explanation" for algorithmic decisions. Historically, Expert Systems researchers have developed explanation methods, and AI systems have garnered attention in the popular press, prompting discussions on AI's implications for social justice and fairness.

This research aims to develop an AI-based disease detection system using machine learning algorithms, promising higher accuracy and efficiency compared to conventional methods. We also plan to employ Feature Mapping and Natural Language Processing to provide clear, text-based explanations for the system's predictions. Given AI's capabilities in processing complex healthcare data, the system can offer accurate diagnoses promptly, potentially saving lives by enabling timely treatments

A. EXPLAINABLE AI

Explainable Artificial Intelligence (XAI) encompasses processes and methodologies that enable human users to understand and trust the outcomes produced by machine learning models. Unlike interpretable AI [1], which focuses on making the model's decisions understandable, XAI goes further by providing explicit explanations for its predictions and actions, ensuring a deeper level of transparency and clarity for users.

B. FOCUS AREAS

We will be focusing on the following:

- Develop an Artificial Intelligence based Disease detection system which can act as a virtual doctor but more efficiently and with a consistent level of accuracy. Developing a medical diagnosis system based on machine learning (ML) algorithms for prediction of any disease can help in a more accurate diagnosis than the conventional method.
- Explaining the result of the Disease Detection system using the Feature Mapping methods and Natural language processing to have a text-based system that can generate the explanation considering the model parameters and understanding.

C. SHAP

SHAP, or SHapley Additive exPlanations, is a prominent method in machine learning designed to enhance model interpretability and facilitate the understanding of model predictions. Its foundation lies in cooperative game theory, particularly Shapley values, which are used to attribute contributions fairly among a group of contributors based on their marginal impacts in all possible combinations. The primary goal of SHAP is to assign each feature's contribution to the prediction of a specific instance [2], thereby providing insights into the factors influencing model output.

The term "additive" in SHAP underscores the property that the explanation for a specific prediction is the sum of individual feature contributions. This additive nature allows the decomposition of model predictions into the impact of each feature, offering a clearer understanding of the model's decision-making process. By adopting Shapley additive explanations, SHAP values provide a unified and comprehensive

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:32:20 UTC from IEEE Xplore. Restrictions apply.

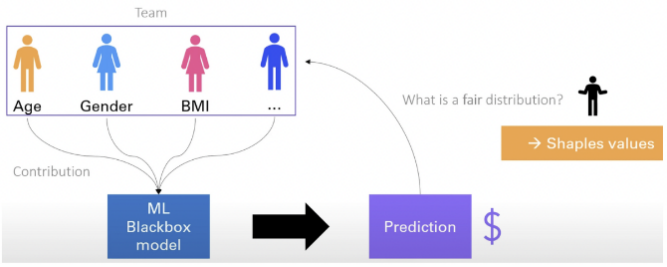


Fig. 1. Shapley Explanation - A Visual Representation of working of Shapley algorithm

measure of feature importance, making it a versatile tool applicable to various machine learning models.

II. Related Work

Numerous studies have explored the use of artificial intelligence techniques for disease detection. Machine learning algorithms have been employed to diagnose various diseases, achieving high levels of accuracy. Similarly, deep learning models have been utilized for the early detection of specific diseases, demonstrating promising results. However, the lack of transparency and interpretability in these models often limits their applicability in clinical settings.

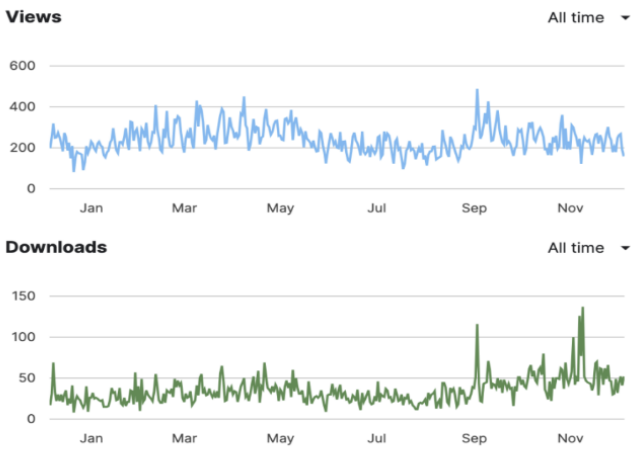
Explainable AI has emerged as a crucial area of research in healthcare to address the interpretability issue of AI models. SHAP (SHapley Additive exPlanations) scores have been utilized to provide interpretable explanations for the predictions made by disease detection models. Another approach involves the use of Large Language Models (LLMs) [3] to generate human-readable explanations, enhancing the transparency of AI-based diagnostic systems.

Another interesting development in XAI is the use of prompting techniques like chain of thought (COT) [4] prompting. Here, the model is prompted to not only generate a prediction but also explain its reasoning process step-by-step. This can provide a more granular understanding of how the model arrives at its conclusions.

III. Methodology

A. DATASET

The dataset is taken from the kaggle named Disease Prediction Using Machine Learning. This dataset provides valuable insights for applying existing knowledge to disease prediction in the medical field. Dataset is in one-hot encoding format which is extracted from the diseases and their respected symptoms showed in different patients. This resulted in 132 unique symptoms. With 132 parameters, the dataset allows for the development of models that can predict 42 different types of diseases. This has the potential to improve physician efficiency and patient care. The dataset consists of two CSV files, one for training and one for testing machine learning models. Each CSV file contains 133 columns.132 of these columns represent various patient symptoms, while the final column indicates the corresponding disease diagnosis. This structure allows models to learn relationships between symptoms and diseases, enabling the classification of 42 different disease types based on a set of symptoms. There are various disease classes included in the dataset like hepatitis, common cold, pneumonia, Tuberculosis, Aids, Migraine, vertigo, Heart Attack, etc. The symptoms include itching, skin rash, red spots, chest pain, bloody stool, joint pains, chills, high fever,stomach pain, dark urine, fluid overload, etc.



The Views and Downloads of the Dataset on kaggle over time.

Fig. 2. Dataset - Views and Downloads

B. NEURAL NETWORK

In our search for the best model architecture to predict the diseases from the symptoms we have tried different machine learning techniques. In our project we have used the custom neural model using the nn module in PyTorch which provides a set of high-level abstractions for building and training neural networks. The F module in PyTorch stands for the functional module, and it provides a set of functions that perform operations without maintaining internal states. PyTorch tensors are the fundamental building blocks for constructing and training neural networks. These are multi-dimensional arrays similar to NumPy arrays but designed to work efficiently on GPUs, allowing for accelerated computation.

In our study we have created this custom neural model in order to predict the disease from 42 predefined classes using the 132 possible symptoms. This model includes two hidden neurons layers, where the first one is double the size of input neurons and the second one is of 300 neurons. We have tried few different layer configurations also and selected this out of all to enhance it's capacity to learn complex patterns in the data and dropped few neurons to prevent over-fitting. The activation function used in each layer is ReLU and there is a final softmax layer to ensure that the output is a valid probability distribution, where each value represents the probability of belonging to a particular class. We have used manual seed 41

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:32:20 UTC from IEEE Xplore. Restrictions apply.

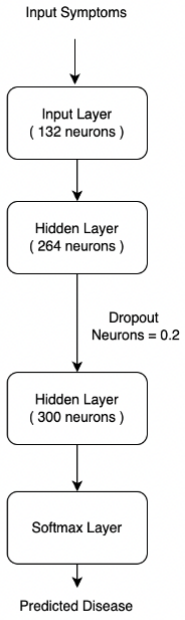


Fig. 3. Flowchart For Custom Neural Network for predicting the disease using symptoms

to ensure result generation is consistent across different runs. In this custom neural network, we have used a cross entropy loss function for multi-class classification. We have used Adam optimizer for our neural network model. Adam is a popular optimization algorithm that adapts the learning rates for each parameter individually. We have trained our custom prediction model up to 100 epochs

C. SHAPLEY VALUES

We have used the SHAP module of python in order to calculate the shapley values associated with the features, which use the game theoretic approach to explain the result of a machine learning model [5].

A set of 500 reference data points were passed to create the kernel explainer object. The explainer object used this subset of training data to estimate how each feature in a new datapoint contributes to the model's prediction. A wrapper was created in order to handle different data type requirements of model and SHAP explainer.

Explainer gives a 2-D vector of size equal to the number of classes to be predicted multiplied by the number of features. In our case it was 41*132. For visualisation we have used the force plot function provided by the shap library itself, which requires the expected values and actual values calculated. Figure (needs to be added) shows one of the generated results.

D. PROMPT DEVELOPMENT

We used an iterative process which is known as prompt engineering [6] in order to create our input to the large language model. During this process, we experimented with several different types of prompts. In each round of prompt engineering, we evaluated the result with BioMistral-7B and ChatGpt-3.5, and selected the most optimum prompt from all. The selected prompt style consisted of 5 parts including context, task, symptoms along with their shapley scores, predicted disease from the custom neural, and line stating the reason explanation from where the explanation would start.

An example is shown below (in figure number 5)

E. REASONING MODEL

In our quest to develop a disease classification system, which can help in providing a better reasoning than the traditional approaches with LLMs, we tried different text generated LLMs available, which can provide reasoning based on the informative prompt we have designed. We first chose the BioMistral language model which is a part of Mistral models family and tailored for the biomedical domain, utilizing the Mistral as its foundation model and further pre-trained on PubMed Central. The model is only available as a research tool and not meant for direct real world application and clinical prediction.

We have used the HuggingFace transformer library to load the BioMistral directly through the pipeline function. Huggingface also has support for loading models directly, which requires to initialise the token and model separately and input needs to be passed through tokenizer in order to generate result from the model. Our pipeline was setted to the text generation task option, so that we can generate the reasoning from the prompt. Prompt and the maximum number of new tokens to be generated were passed for the result generation.

We have also used ChatGpt 3.5 for better result analysis by using prompts we have created along with the Persona Prompting Technique. In this, we have asked Gpt-3.5 to act like a Medical Expert [7] in order to confine its boundary and make it more specific towards Healthcare. Then we have passed the information regarding the input format and their significance into the reasoning so that Gpt can grasp the essence of information in prompt.

We have used two methods to compare the reasoning of BioMistral and GPT-3.5. Method 1 includes the passing of Symptoms along with their shapley values generated from the Shapley explainer and predicted disease given by the custom neural model. The top four symptoms based on the Shapley values were passed for reasoning along with the weights assigned by the Shapley explainer. Method 2 includes the passing of Symptoms with the Predicted Disease by custom neural model.

IV. RESULTS

In our work we have used our custom neural network to predict the disease. The shapley algorithm is used over this model to generate and shapley values for symptoms in each new datapoint. Since the dataset was in one hot encoding, it was easy for neural machines to learn the symptoms - disease relationship and perform the same. The predicted result had a loss of 0.004689. Here is the f1-score results changes for the prediction over 100 epochs.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:32:20 UTC from IEEE Xplore. Restrictions apply.

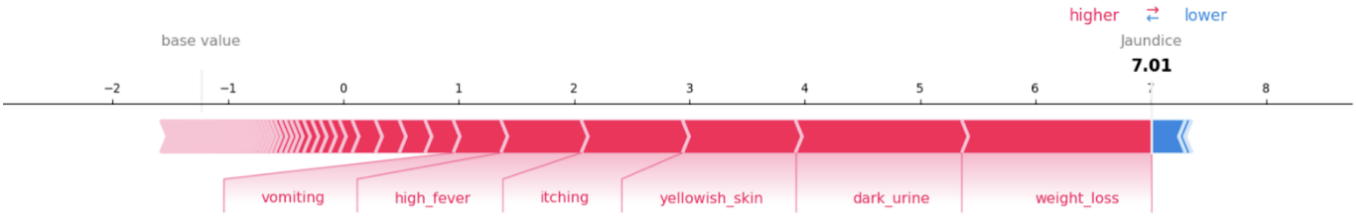


Fig. 4. Shapley Visualization - Example of Visual representation for numeric contribution of each symptoms towards the predicted disease

Context: A disease is predicted based on symptoms and amount of their contribution is represented by scores along with them

Symptoms
1. extra_marital_contacts: 2.6
2. patches_in_throat: 2.49
3. high_fever: 0.94
4. fatigue: 0.22

Predicted Disease: AIDS

Task: Explain the predicted disease in a paragraph considering all symptoms below
Reason Explaination:

Fig. 5. Prompting for Reasoning using LLMs

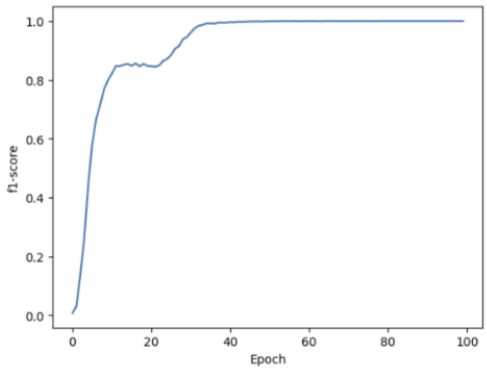


Fig. 6. Change in f1-score over the 100 epochs during training for custom neural prediction

In this study, for the reasoning part we have analyzed the result produced by both the LLM including BioMistral using Hugginface transformer library and ChatGPT-3.5 for Prompt Engineering support.

In the case of BioMistral, the results produced seem to be more satisfactory regarding method 1, which include the Shapley values. Model was able to slightly grab the weightage of symptoms and was able to produce the reason over symptoms having the greater score whereas in case of Method 2, models were mostly mapping the symptoms along with the predicted disease stating that this is the symptoms of predicted disease. Although the overall results were not stable due to the limitation of Model regarding the generation capability and prompt tuning. Due to this the model was not able to generate the complete reasoning including just stating symptoms might be a possible reason.

ChatGpt-3.5 has shown better results over reasoning as it is an advanced LLM with great ability of reasoning generation. We have tried the reasoning generation using both methods. Since the ChatGPT-3.5 is highly developed LLM, it was able to slightly manage without score results too. Using our scoring prompting method, ChatGpt was able to give better weightage to the symptoms that were contributing to the predicted disease. This reasoning includes the identifying specific and non-specific symptoms towards the predicted disease. Giving reason why these symptoms drive the result instead of just mapping the internal cause of symptoms. In this study, we have analyzed a dataset consisting of 50 data samples that were independently reviewed by five different reviewers. The goal was to assess the level of agreement among the reviewers in evaluating the data samples. After careful analysis, we found the agreement score(Fleiss Kappa) among the reviewers to be 50

TABLE I
SURVEY RESULT

Choice	P1	P2	P3	P4	P5	Total
Yes	74%	72%	66%	62%	64%	67.6%
No	26%	28%	34%	38%	36%	32.4%

During our survey, we found an acceptance score of about 70% for our model indicating that most of the reviewers felt positive and confident about this new model of disease detection and explanation.

V. Conclusion

In this research, we pioneered a novel approach to disease detection by leveraging text-based inputs, a shift from the conventional image-based systems. Our emphasis on Explainable AI (XAI) addresses the critical issues of transparency and trust in AI-driven diagnoses, making the decision-making process of the model more understandable and reliable for clinicians and patients alike. By employing a comprehensive and diverse dataset, our model demonstrates real-world applicability across a wide range of health conditions, underscoring its adaptability to complex medical scenarios.

Through meticulous data preprocessing and the implementation of a feed-forward neural network, we achieved commend-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:32:20 UTC from IEEE Xplore. Restrictions apply.

able accuracy in disease prediction. Our study reaffirms that deep learning, particularly when coupled with XAI techniques, offers a promising avenue for enhancing the accuracy and interpretability of disease detection systems. The transition from the inherent opacity of traditional AI to the clarity of XAI transforms the diagnostic process from a 'black box' to a 'white box', offering insights into the model's decision-making process. This not only enhances the credibility of AI-based diagnostic systems but also paves the way for their wider acceptance and integration into mainstream healthcare practices. Our findings highlight the potential of XAI in revolutionizing disease detection, offering a more transparent, trustworthy, and effective alternative to existing AI applications in the medical domain.

VI. Future Work

One promising direction for future research is the integration of multimodal data, such as medical images, patient history, and laboratory test results, along with textual symptom descriptions. Incorporating such diverse data types could provide a more holistic view of the patient's condition, thereby improving the accuracy and reliability of disease classification. Advanced techniques such as multi-modal learning and attention mechanisms can be employed to effectively fuse and utilize these different data sources.

Moving beyond disease classification, future research can focus on developing a more comprehensive AI system that not only identifies diseases based on symptoms but also provides personalized treatment recommendations [8]. The AI systems should not only try to map to symptoms with cause, but should also have knowledge of which symptoms are deriving the result and hold how much weightage. This can be achieved by integrating the AI model with the algorithms which can explain the model results such as lime and shap as we do but in a much more efficient way and use medical knowledge bases, drug databases, and treatment guidelines. By considering individual patient characteristics such as age, gender, medical history, and genetic information, the AI system can generate tailored treatment plans and recommendations. Reinforcement Learning techniques can be employed to continuously learn and adapt the treatment recommendations based on the patient's response and feedback.

Lastly, as AI technologies continue to advance and become increasingly integrated into healthcare systems, it is crucial to address the ethical and regulatory challenges associated with their implementation. Future research should focus on developing robust frameworks and guidelines to ensure the responsible and ethical use of AI in healthcare. This includes addressing issues related to data privacy and security, transparency, interpretability and explainability of AI algorithms, and the potential biases and limitations inherent in AI-based diagnostic systems. Collaborative efforts involving healthcare professionals, ethicists, policymakers, and technologists are essential to establish a comprehensive regulatory framework that promotes the safe, equitable, and effective use of AI in healthcare while maintaining patient trust and confidence in the healthcare system.

References

[1] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1135-1144, 2016.

[2] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.

[3] O. Ben-Shoham and N. Rappoport, "CPLLM: Clinical prediction with large language models," *arXiv preprint arXiv:2309.11295*, 2023.

[4] S. Hao, Y. Gu, H. Ma, J. J. Hong, Z. Wang, D. Z. Wang, Z. Hu, "Reasoning with language model is planning with world model," *arXiv preprint arXiv:2305.14992*, 2023.

[5] Erik Štrumbelj and Igor Kononenko. "Explaining prediction models and individual predictions with feature contributions". In: *Knowledge and information systems* 41.3 (2014), pp. 647–665.

[6] Wang, J. et al. Prompt engineering for healthcare: methodologies and applications. Preprint at https://doi.org/10.48550/arXiv.2304.14670 (2023).

[7] M. Krusche, J. Callhoff, J. Knitza, and N. Rufer, "Diagnostic accuracy of a large language model in rheumatology: comparison of physician and ChatGPT-4," *Rheumatology International*, vol. 44, no. 2, pp. 303-306, 2024. doi: 10.1007/s00296-023-05464-6

[8] M. Jin, Q. Yu, C. Zhang, D. Shu, S. Zhu, M. Du, Y. Zhang, and Y. Meng, "Health-LLM: Personalized retrieval-augmented disease prediction model," *arXiv preprint arXiv:2402.00746*, 2024.

[9] Y. Kumar, A. Koul, R. Singla, and M. F. Ijaz, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda," in *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 7, pp. 8459-8486, 2023.

[10] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138-52160, 2018.

[11] T. Savage, A. Nayak, R. Gallo, E. Rangan, and J. H. Chen, "Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine," *npj Digital Medicine*, vol. 7, no. 1, p. 20, 2024, doi: 10.1038/s41746-024-01010-1.

[12] K. Gaurav, A. Kumar, P. Singh, A. Kumari, M. Kasar, T. Suryawanshi, "Human Disease Prediction using Machine Learning Techniques and Real-Life Parameters," *International Journal of Engineering, Transactions C: Aspects*, vol. 36, no. 06, pp. 1092-1098, 2023.

[13] Thirunavukarasu, A. J. et al. Large language models in medicine. *Nat. Med.* 29,1–11 (2023).

[14] S. M. Bsharat, A. Myrzakhan, and Z. Shen, "Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4," *arXiv preprint arXiv:2312.16171v2*, 2024.

[15] Ibrahim, I. and Abdulazeez, A., "The role of machine learning algorithms for diagnosing diseases", *Journal of Applied Science and Technology Trends*, Vol. 2, No. 01, (2021), pp. 10-19. doi:10.38094/jastt20179.

[16] Ren, Q., Cheng, H. and Han, H., "Research on machine learning framework based on random forest algorithm," in *AIP conference proceedings*, AIP Publishing LLC. Vol. 1820, (2017), 080020.

[17] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez,Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, and others. 2019. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. *Information Fusion* (2019).

[18] Diogo V Carvalho, Eduardo M Pereira, and Jaime S Cardoso. 2019. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* 8, 8(2019), 832.

[19] Shahadat Uddin, Arif Khan, Md Ekramul Hossain, and Mohammad Ali Moni. 2019. Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1):1–16.

[20] Jain D, Singh V (2018) Feature selection and classification systems for chronic disease prediction: a review. *Egypt Inform J* 19:179–189.

[21] S. Vijayarani, S. Dhayanand, "Liver disease prediction using svm and naive bayes algorithms," *International Journal of Science, Engineering and Technology Research (IJSETR)*, 4(4), 816 (2015).

[22] Singhal, K. et al. Large language models encode clinical knowledge. *Nature* 620,172–180 (2023).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:32:20 UTC from IEEE Xplore. Restrictions apply.

[23] Nori, H., King, N., McKinney, S. M., Carignan, D. & Horvitz, E. Capabilities of GPT-4 on medical challenge problems. *arXiv preprint arXiv:2303.13375*(2023).

[24] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. "Huggingface's trans-formers: State-of-the-art natural language processing". In: *arXiv preprint arXiv:1910.03771*(2019).

[25] Q. V. Liao, D. Gruen, and S. Miller, "Questioning the AI: Informing design practices for explainable AI user experiences," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1-13, 2020, doi: 10.1145/3313831.3376590.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:32:20 UTC from IEEE Xplore. Restrictions apply.

Check for updates

Received: 17 November 2023 | Revised: 17 May 2024 | Accepted: 19 June 2024

DOI: 10.1002/fam.3229

RESEARCH ARTICLE

WILEY

Exploring the potential of dual-metallized PET towards improving the efficiency of outermost reflective layer in fire proximity clothing

Shivangi Dwivedi[1,2] | Richa Srivastava[2] | Prasun Kumar Roy[1]

[1]Centre for Fire, Explosive and Environment Safety, Defence Research and Development Organization, Delhi, India

[2]Department of Applied Chemistry, Delhi Technological University, Delhi, India

Correspondence
Prasun Kumar Roy, Centre for Fire, Explosive and Environment Safety, Defence Research and Development Organization, Delhi, India.
Email: pkroy.cfees@gov.in

Abstract

The outermost layer of a fire proximity suit needs to conform to a strict requirement of radiant protection performance (RPP) ≥ 20s, which is indicative of its ability of offering a protection for at least 20s duration from second degree burn upon radiant heat exposure (84 kW/m^2). Typically, this layer is fabricated by laminating a single-side metallized PET (SMPET) layer with glass fabric. However, upon erosion of the deposited metal, this laminate is rendered unsuitable due to loss of reflectivity. Here, we explore the possibility of replacing the SMPET with its dual-metallized analogue (DMPET) and determine the effect of increasing the optical density (OD) on the adherence and protection level. Metallized films with OD varying from 2.2 to 4.8 were laminated with glass fabrics of twill, satin and plain weave pattern using a silicone adhesive. The peel adhesion strength of laminates prepared using DMPET was found to be higher (1.01 ± 0.03 N/mm), as compared to SMPET (0.63 ± 0.03 N/mm) and the resulting films did not undergo delamination during flexing. Laminates prepared from satin woven glass fabric exhibited lowest flexural rigidity followed by twill and plain woven glass fabric. Protection offered by the laminate from convective heat was quantified in terms of the thermal protective performance (TPP), and the abraded laminate prepared using DMPET (OD-4.8) was found to meet all the mandatory requirements of proximity clothing, offering an RPP of 27 s and a TPP of 62 cal/cm^2 s. In comparison, SMPET laminates exhibited lower level of adhesion and offered an RPP of only 7.5 s.

KEYWORDS
fire proximity, laminate, reflective layer, thermal stability

1 | INTRODUCTION

Depending upon the fire scenario, fire-fighting operations are primarily categorized into three different categories namely structural, wildland and proximity.[1–4] Amongst these, proximity fire-fighting is a specialized operation which involves activities associated with the highest level of conductive, convective and radiative heat exposure.[5] Fire protective clothing is required to be worn by the fire-fighters to guarantee requisite level of thermal protection to the personnel involved in these operations. A proximity firefighter clothing is assembled out of several layers, each having its own specific role to play as per its placement in the configuration (Figure S1). Since very high level of radiant heat exposure is expected during proximity operations, the outer shell is expected to reflect at least 90% of the radiant heat.[4] This is quantified in terms of radiant protective performance (RPP), where in the layer is initially abraded and subsequently subjected to extremely high levels of radiant heat flux as observed under flash fire conditions (84 kW/m^2). Conventionally, the reflective layer is

 © 2024 John Wiley & Sons Ltd.

prepared by laminating an aluminized polyester film over an underlying fabric.[5,6] Unfortunately, a significant fraction of the topmost metal coating gets eroded during abrasion; therefore, the requisite level of radiant protection is not achieved. In addition, the layer is expected to maintain its integrity by not undergoing any damage in terms of cracking or delamination during flexing cycles, even under wet conditions. This essentially implies that there should be excellent adhesion of the deposited aluminium with the PET film as well as that of the resulting metallized film with the base fabric.

Aluminized fabric laminates have been reported to reflect 80% to 90%[7,8] of radiant heat, which is quantified in terms of percentage reflectance and radiant protection levels. Earlier, aluminium foil was used for preparing aluminized fabric laminates; however, the tear resistance of the foil is rather low.[9] With technological advancements, metallized PET films have become more common, as they offer better mechanical properties at reduced costs. In the wake of these developments, metallized PET has practically replaced aluminium foil not only in the field of food packaging but also in firefighter clothing as well.[10]

Recently, the preparation of reflective aramid fabric laminate with both double- and single-sided aluminized films has been reported by Jin et al.[7] The entire assembly comprising of the developed reflective layer, moisture barrier layer and thermal barrier layer was subjected to a radiative flux of 40 kW/m^2, and the configuration containing dual-metallized films was found to offer high radiant protection ($RHTI_{24}$ > 100s); however, it was not sufficient to meet the criteria mandated in EN 1486 ($RHTI_{24}$ > 120s).[11] Furthermore, the effect of higher radiant flux (84 kW/m^2 as mandated in NFPA 1971) on the protection level offered by the laminate was not studied.[12]

Liu et.al. have reported the use of pigment coating technique for the preparation of aluminized fabric, where aluminium pigment was formulated with intumescent flame-retardant system in thermoplastic polyurethane solution for coating aramid fabric.[8] The commercial aluminized fabric laminate was found to display excellent thermal insulation performance and exposures to heat fluxes (21 kW/m^2) did not result in first or second degree burns even after duration of 30 s.

In our recent study, we employed a single-sided aluminized PET film for the fabrication of a reflective laminate meant for fire proximity operations. However, to achieve the radiant performance requirements, the adhesive was formulated with aluminium powder,[13] which led to reduced flexibility.

We hypothesize that by increasing the optical density (OD) of the metallized PET, it is possible to meet the stringent requirements of radiant protection mandated by NFPA 1971.[12] Furthermore, by replacing the single-metallized PET film with its dual-metallized analogue, it is possible to overcome the damage caused by abrasion. In the present work, we prepare laminates of single- and dual-metallized PET films with glass substrates of different weave patterns. Of particular interest was to explore the effect of increasing the OD of the dual-metallized PET films on various properties of the laminate, like adhesion, protection level against convective and radiative exposures as well as flexibility under both dry and wet conditions, for establishing its candidature as a material as an outer layer of fire proximity suits.

2 | EXPERIMENTAL

2.1 | Materials

Metallized biaxially oriented PET film (single side aluminized and double side aluminized) were obtained from SRF, Indore. Glass fabrics of three different weaves (satin weave, plain weave and twill weave, 390 gsm, thickness ~ 0.41 mm) were obtained from Montex glassfiber industries. A poly(siloxane)-based resin (SE 9152 HT) was procured from Dow and used as a laminating adhesive. Neoprene-coated glass fabric (120 gsm) was obtained from Raksha Polycoat, Needle-punched Oxidized PAN fabric (200 gsm) was obtained from Welspun, and a twill woven fabric (165 gsm, 50:50 FR Viscose- m-aramid fabric) was obtained from Shingora Textiles, Ludhiana for use as an ensemble for TPP testing.

2.2 | Metallization

The metallization of PET was performed on a vacuum metallizer unit (General Vacuum Equipment). In brief, the metal to be deposited (aluminium) was heated to its melting point in a vacuum chamber and subsequently condensed over biaxially oriented polyester films (BOPET, 14 μ). For improving the adhesion, the surface of BOPET was coated with a thin layer of amorphous copolyester. A representative schematic is presented in Figure S2. The stretched film was forced to travel at high speeds over the aluminium evaporators, where it received both the radiated heat and the latent heat from the condensing aluminium. The required OD was achieved by controlling the speed of the moving web. The metallic vapour condenses over the film, as it proceeds to the cooling roll, following which it is rewound and removed as a reel. For preparing dual-side coated films, the operation was performed twice, once on each side. The sample designations of the films are presented in Table 1.

2.3 | Preparation of reflective laminates

A moisture curable poly(siloxane) adhesive was used to laminate the metallized BOPET film with the underlying glass substrate, using a hand layup technique, while maintaining an adhesive thickness of ~0.04–0.08 mm. The laminates were left undisturbed under room temperature conditions for 24 h to ensure complete curing. The designation and grammage of the laminates prepared have been presented in Table S1.

2.3.1 | Characterization technique

Water vapour transmission rate

The water vapour transmission rate (WVTR) of the metallized BOPET films was determined according to procedure B (upright cup method) of ASTM E 96.[14] Circular specimen (6 no's, 5 cm diameter) was cut

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Details of the metallized films and their sample designation.

Sample designation	Metallization	Optical density	Grammage (gsm)	Metal bond strength (kg/25 mm)
SMPET-2.2	Single side	2.2 ± 0.1	17.0	>1.3
DMPET-3.1	Both sides	3.1 ± 0.1	17.0	>1.3
DMPET-4.8	Both sides	4.8 ± 0.4	17.1	>1.3

and placed on aluminium cup (155 mL capacity), filled with 100 mL of distilled water. The cup was clamped to position by covering with a gasket, and the assembly was placed under controlled temperature and humidity conditions (23 ± 1°C, 50 ± 2 % RH). The weight of each cup was recorded at regular intervals, and the WVTR (gh/m^2) was quantified as follows:

$$\mathrm{WVTR} = \frac{G}{t \times A} \quad (1)$$

where G refers to mass change (g), t being the time associated with the mass loss (h), and A refers to the test area (m^2). The data have been reported in g/24 h/m^2 along with the standard deviation.

Fourier transform infrared spectroscopy (FTIR)

The structural characterization of poly(siloxane) was performed using Thermoscientific FTIR (NICOLET 8700) analyser with an attenuated total reflectance (ATR) crystal accessory in the wavelength range 400–4000 cm^{-1}.

Thermo-gravimetric analysis

The thermal decomposition behaviour was analysed using Perkin Elmer Diamond STG-DTA over the temperature range 50–600°C.

Contact angle studies

The contact angle and surface free energy of the metallized BOPET sample were determined using a Drop Shape Analyser (DSA100E) from KRÜSS Scientific. For contact angle measurements, distilled water was used as the probing liquid, and for surface energy measurements, contact angle measurements were performed with two probing liquids (water and N, N-DMF).

Waterproofness

The degree of waterproofness offered by the metallized film was quantified using Hydrostatic Head Tester as per ISO 811. In brief, one of the face of the specimen was subjected to a steadily increasing hydrostatic pressure (10 mbar/min) till penetration was observed at three different sites on the opposite face.

2.3.2 | Characterization of laminate

Dimensional stability

In view of the targeted application of the laminate, the dimensional stability of the fabric in terms of resistance to thermal shrinkage was determined as per ASTM F2894[15] at a temperature of 260°C. For this purpose, a sample of standard dimensions (381 × 381 mm^2) was suspended through a metal hook at a temperature of 260 ± 3°C for 5 min and examined for evidence of ignition, melting, dripping or separation and resultant dimensional change was recorded.

Physical properties of fabrics

The stiffness of fabric was determined as per ASTM D1388[16] using a Shirley stiffness tester (Figure S3). Prior to the tests, the samples were preconditioned at 25°C and 65% relative humidity (RH) for 24 h. Five replicates were tested in both warp and weft directions. The bending rigidity was calculated as follows:

$$G = W \times c^3 \quad (2)$$

where G is the fabric's bending rigidity (mg cm), W is the weight per unit area (mg/cm^2), and the bending length c is half the length of the overhang (L, cm). The overall bending rigidity was determined as follows:

$$G_0 = \sqrt{G_w \times G_f} \quad (3)$$

where G_o is the fabric's overall bending rigidity (mg cm), and G_w and G_f refer to the bending rigidity (mg cm) in the warp and weft directions, respectively.

Dynamic mechanical properties

The dynamic mechanical analysis (DMA) was conducted on rectangular samples (10 × 4 × 0.8 mm^3) at a fixed frequency of 1 Hz, temperature range of 20–180°C at a heating rate of 10°C/min and in strain sweep mode using a dynamic mechanical analyzer (Netzsch, Artemis-242E) and tensile fixture.

Rheology

Anton Paar Rheometer MCR-102 has been used to study the rheology associated with the curing of the siloxane resin. All the experiments were performed using a parallel plate accessory (diameter: 25 mm) with a gap of thickness 0.5 mm. Temperature-sweep experiments were performed in oscillatory mode over the temperature range 30–200°C at 5°C/min. A constant strain of 0.5% and constant angular frequency of 10 rad/s was maintained during each interval.

Peel adhesion strength

The peel adhesion strength (N/mm) was established as per ASTM D1876[17] as both the substrates were flexible. For this purpose,

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

laminates (317 mm × 25 mm) were cut, where 241mm of the region was kept bonded, with the remaining 76 mm being non-bonded. The non-bonded region was fixed within the grips of a tensile testing equipment and subjected to a strain rate of 10 mm/min. The peel adhesion strength (N/mm) was determined over at least a length of 200 mm of the bond line after the initial peak was observed. A total of six replicates were tested, and the average of the adhesion strength values with standard deviation has been reported.

Wet flex test

For this purpose, the samples were cut into 20 cm × 10 cm rectangular-shaped specimens in warp direction, which were then placed in a beaker (1000 mL) at 60 ± 2.8°C for 15 min. Blotted dry samples were flexed at a rate of 45 cycles per minute, using a motion that repeatedly twists the film. The samples were subjected to 1000 full flex cycles on the Flex Tester in accordance with standard test method as per NFPA 1971.[12]

Vertical flammability tests

Vertical flame resistance test was conducted on a flammability tester (Prolific Engineers, India), as per ASTM D6413-99.[18] In brief, the sample of requisite dimensions (5 cm × 25 cm) was exposed to a propane flame of 36 mm height for a period of 12 s and the flame time (time for which the material continues to burn after flame removal), after glow time (time associated with the material glow post-flame extinguishment) and char length (length of the area damaged) were recorded.

Heat resistance and thermal shrinkage

The dimensional stability of the laminate in terms of resistance to heat and thermal shrinkage was determined as per ASTM F2894[15] at a temperature of 260°C. For this purpose, sample of standard dimensions (381 × 381 mm^2) was suspended by a metal hook in an oven maintained at the mentioned temperature for 5 min and examined for evidence of ignition, melting, dripping or separation and resultant dimensional change.

Scanning electron microscopy (SEM)

ZEISS EVO 50 series instrument, operating at 10 kV acceleration voltage, was used to study the char observed in the reflective layer after the testing of thermal protective performance (TPP). In addition, the satin weave pattern of the glass fabric utilized in the present study was analysed. Aluminization on both sides of the aluminized polyester film was confirmed by EDAX mapping. For the investigation, the specimens were mounted on the aluminium sample stub using double sided carbon tape and were subjected to gold coating prior to the analysis.

Heat transmission on exposure to radiant heat

For this study, a bench-top RPP tester was used, which utilizes a bank of 9 quartz lamps to deliver a total radiant heat flux of 84 kW/m^2 (as specified in NFPA 1971)[12] on the reflective surface of the developed laminates. A schematic of the experimental setup used is shown in Figure S3. Specimens (3 inches × 10 inches) were cut and abraded using an oscillatory cylindrical abrasion tester. In brief, the specimens were mounted on the oscillating drum of the apparatus, where a hard textured cotton duck abradant was employed to impart 300 cycles of abrasion under a load of 1.36 kg (3 lb). The abraded samples were then placed in the sample holder of the RPP tester, which was positioned in front of the quartz tubes and blocked by a shutter. A copper calorimeter was placed behind the sample for recording the temperature rise as a function of time.

Heat transmission on exposure to flame

For these studies, the developed reflective laminate, vapour barrier (neoprene-coated glass fabric), a thermal barrier based on spun lace OPAN followed by an inner liner (50:50 FR Viscose- m-aramid fabric), with dimensions 6 inches × 6 inches was assembled and subjected to a predetermined heat flux as per ASTM D 4108.[19] The flow rates of propane and air were adjusted to generate a heat flux of 84 kW/m^2, and a copper calorimeter was placed at the back of the multilayer component assembly. The temperature rise as a function of time was plotted alongside the Stoll curve (second degree), to arrive at the time associated with second degree burn which in turn was used to determine the TPP as the product of the incident heat flux with the intersect time recorded.

3 | RESULTS AND DISCUSSION

The present work aims at improving the performance of the reflective outermost layer of fire proximity clothing by replacing the single-metallized BOPET-glass laminate with its dual-metallized analogue. The effect of increasing the OD of the MPET film on the radiant protection level has also been studied.

3.1 | Metallized BOPET

Metal-PET adhesion is an extremely important parameter for the present application as the layer needs to maintain its reflectivity and integrity even under extreme conditions, especially high temperatures and high radiant flux. It is to be noted that the adhesion of metal with the substrate is governed by several factors, particularly the nature of the surface and its interaction with the incoming metal vapours. The adhesion of metal with PET can be improved by either introducing functionalities on the underlying substrate, commonly known as chemical bonding, or by increasing the mechanical interlocking between the two. In the present work, PET was first coated with an amorphous aliphatic copolyester to increase the degree of mechanical interlocking.[20] By varying the coating thickness, which was in turn controlled by altering the speed of the PET film over the evaporator unit, different optical densities could be obtained. Characteristic physical properties of the metallized BOPET films are presented in Table 2.

In line with previous studies, the metallized films exhibited excellent resistance to the transfer of water vapour, which reflected in

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Physical properties of metallized BOPET.

Sample designation	Optical density	Grammage (gsm)	Thickness (μ)	Waterproofness (mbar)	Metal bond strength (kg/25 mm)	WVTR (g/m^2/day) (23°C, 50%RH)
SMPET-2.2	2.2 ± 0.1	17.0	14	940	>1.3	4.2
DMPET-3.1	3.1 ± 0.1	17.0	14	941	>1.3	4.8
DMPET-4.8	4.8 ± 0.4	17.1	14	943	>1.3	4.8

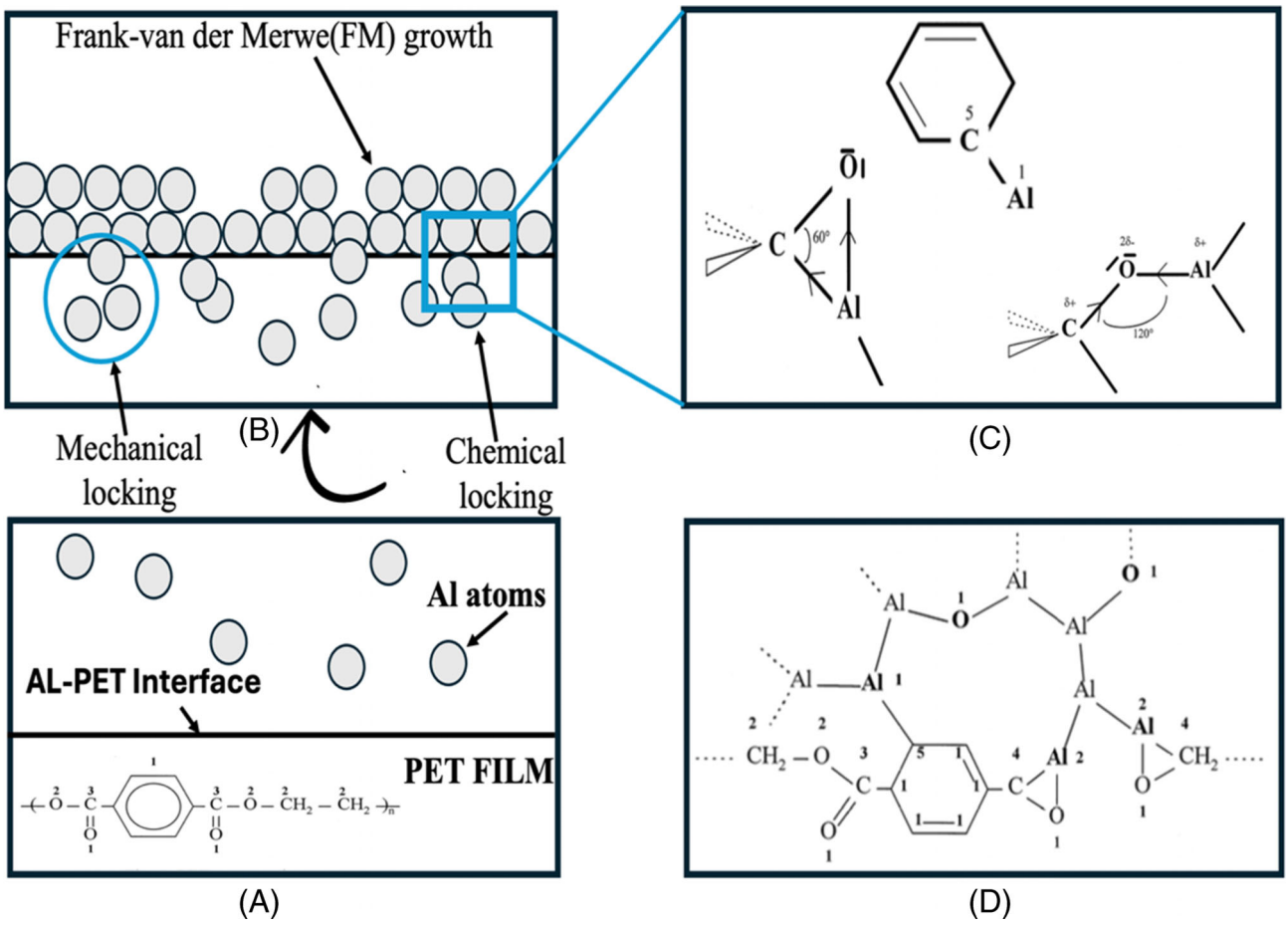


FIGURE 1 Schematic of the metallization process (A) deposition of Al on the interface, (B) mechanical and chemical locking, (C) species identified through XPS and (D) representative model of Al-PET interface.

terms of extremely low WVTR values. Additionally, the films also possess high degree of waterproofness, due to which they find application in food packaging.[21] As expected, dual metallization led to a significant increase in OD, which is indicative of the light blocking ability of the film.

In the past, extensive studies have been performed on the Al-PET interface to understand the underlying pathway behind the metallization process. Most of these studies are based on X-ray photo-electron spectroscopy (XPS) which are aimed towards identification of the chemical bonds at the interface.[22–24] A general schematic of the metallization process is presented in Figure 1. The chemical structure of the PET has also been included in the figure for the ease of visualization. It can be seen that the repeat unit has 10 C, 4 O and 8 H atoms, which have been labelled in the figure to distinguish between the different sites (Figure 1A). The chemical species identified through XPS and a simplistic model of Al-PET surface is also included in the Figure 1C. As the Al atoms strike the polymer surface, they tend to lose their energy, thereby penetrating into the amorphous regions of the polymer resulting in 'Mechanical locking'. In addition, chelation occurs between the surface functionalities and the metal, a phenomenon referred to as 'Chemical locking' (Figure 1B). It is the interplay of both these phenomenon, which define the overall adhesion levels observed in the metallized film. XPS studies suggest that the incoming Al interacts with the C=O and C–O groups on the PET surface to form an oxy–carbide complex, Al–O–C, in which the Al is bonded to both C and O (Figure 1C).[25] Al has been reported to form bonds with three fourths of the C=O groups and to half of the C–O groups

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

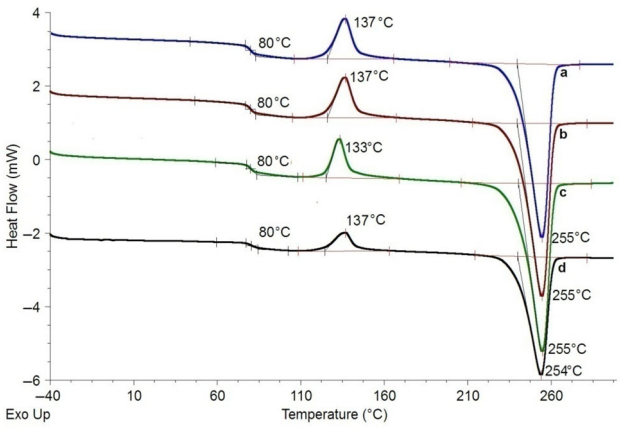


FIGURE 2 DSC trace of PET and metallized PET films (A) DMPET-4.8, (B) DMPET-3.1, (C) SMPET-2.2 and (D) PET.

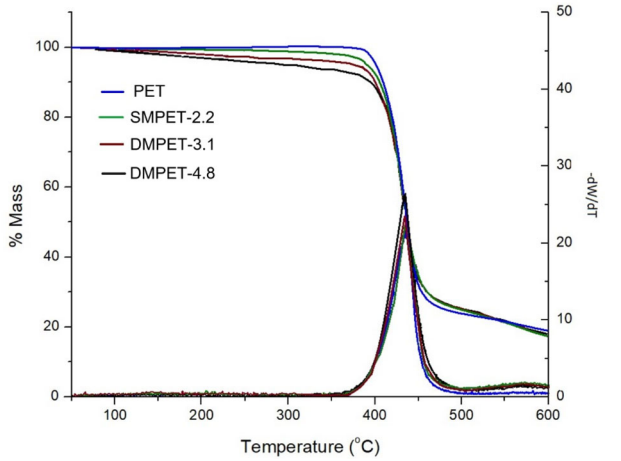


FIGURE 3 TG-DTG trace of PET and metallized PET.

present on the surface. Carbide-like bonds (Al–C) have also been observed in the interface, the presence of which has been attributed to the reaction of Al with the surface phenyl moieties, resulting in the loss of their aromaticity and thus in the disappearance of the C(1s) $\prod$–$\prod^*$ transition peak.[23] However, it is to be noted that there is still ambiguity in this structure, as the valency of the Al remains uncertain, and further studies are required for proper elucidation. A simplistic model of the Al-PET interface as suggested by Sandrin et.al[24] is also included in Figure 2D. The metal atoms continue to grow layer by layer into a continuous film, initially following the Frank-van der Merwe growth mode.[26] However, after formation of a few monolayers the growth pattern shifts to a mixed mode (Stranski-Krastonov).[27] Detailed studies have revealed that the increased mechanical penetration of the metal into the amorphous copolyester is the primary reason behind the high adhesion levels in the coated film as the chemical nature of the copolyester and PET is rather similar. According to TEM analysis of metallized chemically coated PET film, the depth of penetration is ~50–60Å in comparison with ~30–40Å for uncoated film.[23,28]

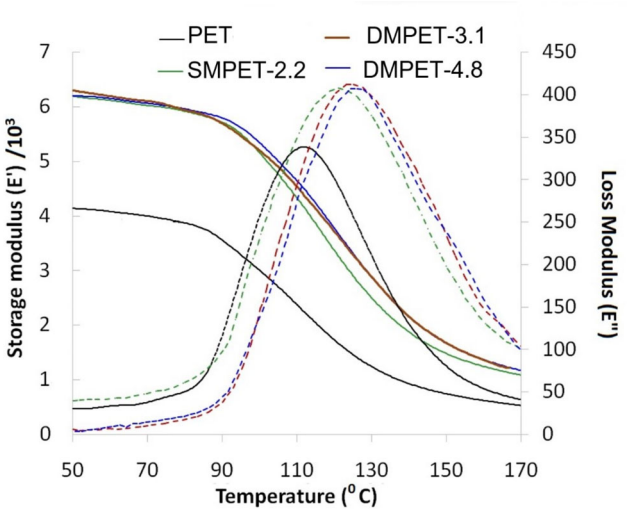


FIGURE 4 Variation of storage modulus (solid lines) and loss modulus (dashed lines) with temperature.

3.2 | Thermal properties

The DSC traces associated with PET and MPET are presented in Figure 2. Neat PET exhibits a glass transition temperature of ~80°C, followed by crystallization at ~136°C, and final melting at ~254°C. Metallization was not found to affect these properties appreciably, which remained within a range of ±1°C.

The TG-DTG traces of PET, single- and dual-metallized PET of different optical densities are presented in Figure 3. A single-step decomposition was evidenced in all the traces. It is interesting to note a slightly larger mass loss in the initial periods of heating in the metallized films, which is reflected in terms of lower T5% (Table S1). However, the temperature associated with the subsequent major mass loss step (T_{max}) remained unaffected. The mass loss in the initial stages can be attributed to the presence of the amorphous copolyester with which the PET film is coated in order to improve its adhesion level with the metal. The copolyester being amorphous permits the mechanical penetration of the incoming aluminium into the film (mechanical locking), thereby increasing the adhesion levels. The copolyester exhibits relatively lower thermal stability than the underlying PET and therefore accounts for the higher mass loss in the coated samples. Furthermore, the coating had to be performed on single and both sides of the film for preparing SMPET and DMPET, respectively, which accounts for the lower T5% in the dual-metallized films.

3.3 | Dynamic mechanical properties

Variation of characteristic properties namely storage modulus and loss modulus of BOPET as well as MPET films due to increasing temperature is presented in Figure 4. As expected, the storage modulus 'E' decreases with temperature due to softening of the polymer, which is particularly pronounced in the region ~90–150°C, where the glass transition phenomenon was evident. This region is also associated

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

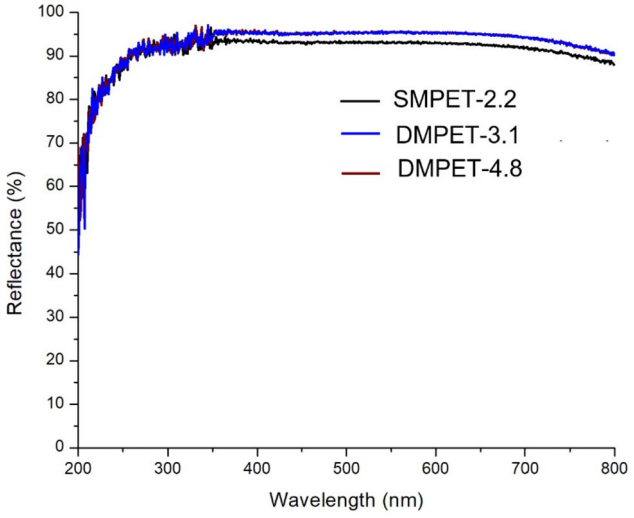


FIGURE 5 Reflectance of metallized film as a function of wavelength.

with a characteristic peak in the loss modulus trace. In line with previous studies, the storage modulus of MPET is relatively higher than the neat films.[28] This can be attributed to the deposition of the metal on the polymer, which restricts the free motion of main chain in various directions, resulting in higher stiffness. In addition, metallization brings a slight shift in the temperature associated with the glass transition phenomenon. The T_g as determined from the peak of the tan δ trace was found to be 124°C for neat BOPET film, which increased to ~138°C for MPET (single metallized), 139°C for MPET (low-OD dual metallized) and 140°C for high-OD dual-metallized films.

3.4 | Reflectivity

The reflectance spectra of both types of MPET over the wavelength region 200–800 nm are presented in Figure 5. It can be seen that the reflectance exceeded 95% over this entire region and as expected the dual-metallized PET films exhibited higher reflectance as compared to the single-metallized films. In comparison, previous studies reveal that coloured textiles are capable of reflecting only ~10% of the radiation.[4] In this direction, Quintiere et al.[29] studied a variety of textile samples for thermal protection from fires, including cotton, nylon and aromatic polyamides of different colour (white, yellow, black and silver). It was noted that the colour of the textile samples had an insignificant influence on reducing the radiative heat load, and more interestingly, the influence of colour was negated post-wetting. Subsequent studies have revealed that construction of aluminized fabrics via pigment coating method are highly effective in reducing the radiative heat load on the textile sample, and this has been attributed to the increase in spectral reflectance due to aluminium coating.[30] Furthermore, studies have also been performed on aluminium foil laminated basalt and glass fabrics. Due to the differences in the yarn parameters and weave structures, their spectral reflectance was found to differ significantly.[31] As expected, the TPP of aluminized fabrics was found to be better than those of the substrate fabrics.

3.5 | Contact angle measurements

The contact angle and the surface energy of the metallized and non-metallized face of the film are presented in Figure S5. Due to the polar nature of the glass fabric substrate, it was not possible to determine the contact angle as they completely absorbed the water and DMF droplets. The contact angle associated with the non-metallized face of the single-metallized PET film was 79 ± 0.11°. Post metallization, a decrease in the contact angle was observed, with both the faces of the dual-metallized film exhibiting contact angles of the same order (57 ± 0.11°) with a surface energy of 43 ± 0.82°, indicative of the relatively higher hydrophilicity of the film.

3.6 | Rheology

For all practical applications, the adhesive resin is coated on the fabric, prior to lamination with the MPET film. For obtaining a uniform coating, the viscosity of the formulation should be in the range of ~2–4 Pa.s, and additionally, sufficient time also must be ensured (15–20 min) for curing.[32] Therefore, it is essential to study the rheological changes to establish a suitable 'processing window'. During any polymerization reaction such as the curing of poly(siloxane) at a given temperature, the storage and loss moduli evolve over time, providing useful information about the polymerization kinetics and the viscosity of the system.[33] The follow-up of their evolution as a function of time is referred to as rheo-kinetics. Figure 6 shows the evolution of storage and loss moduli as a function of time at different temperatures (30–60°C), which was used to arrive at the activation energy associated with curing. During the initial stages, the poly(siloxane) resin is a liquid with viscosity ~10.7 Pa.s and storage modulus ~25Pa. Due to moisture curing, an increase in viscosity and storage modulus was observed. The time associated with the crossover of storage and loss modulus is referred to as the 'gel time', which basically gives an indication of the time available for processing.[31]

The variation of gelation time with temperature was used to arrive at the activation energy associated with the curing process. The variation of t_{gel} as a function of absolute temperature and the derived activation energy is presented in Figure S6. The activation energy was determined to be 43.53 kJ/mol, which appears to be suitable for commercially viable adhesive formulations.[34]

3.7 | Laminate of glass fabric with dual-metallized PET

Previously, different adhesives have been explored for lamination, for example, fluorinated elastomers, polyurethane adhesives and rubber-based adhesives. Silicone-based elastomers exhibit high

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

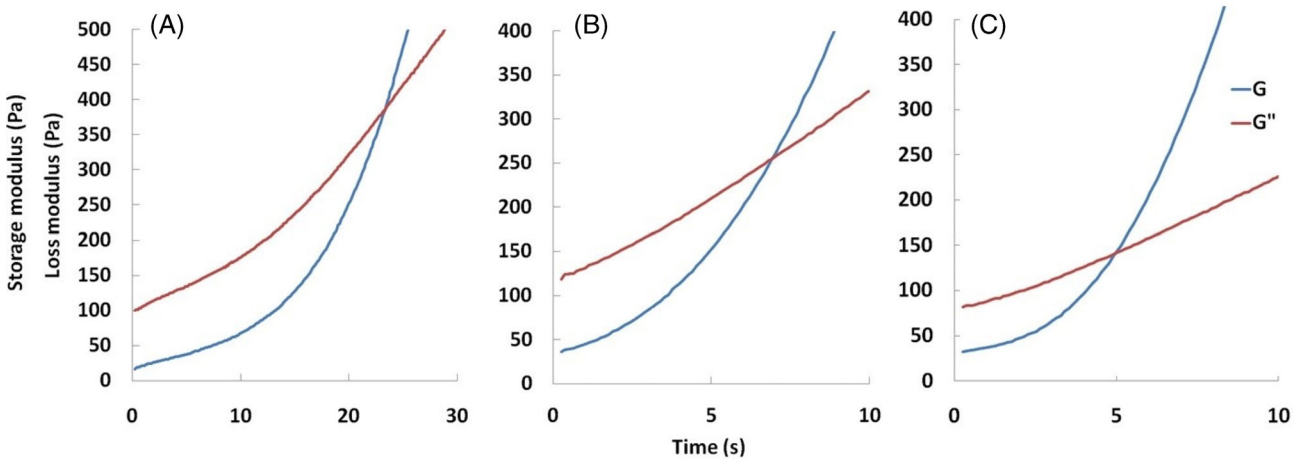


FIGURE 6 Evolution of storage and loss moduli as a function of time at different temperatures (A) 30°C, (B) 45°C and (C) 60°C.

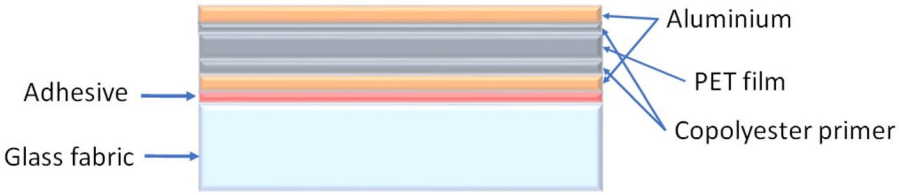


FIGURE 7 A schematic illustration of the laminate components.

thermal stability and high flexibility[35]; therefore, the same was chosen as the adhesive for the said laminate. In our previous work, we have reported an epoxy-based adhesive to laminate a single-side metallized PET (SMPET) with glass fabric. The laminate met all the requirements; however, it was relatively inflexible in view of the nature of the epoxy-based adhesive used for lamination purposes.[13] To address this issue, a poly(siloxane) resin has been employed in the present work. The moisture curable resin contains alkoxy or acetoxy groups, which undergoes hydrolysis in the presence of the atmospheric water vapour, thereby forming silanol (Si–OH) functionalities. These in turn undergo condensation to form a cross linked network, thereby releasing water molecules, as depicted in a representative reaction in Figure S7.

The DSC curve of the cured resin is presented in Figure S8. A glass transition temperature of −49°C was observed, which is responsible for its flexible nature.[36] The poly(siloxane) films exhibited the requisite level of flame retardancy as confirmed through vertical flammability tests. The FTIR of the cured poly(siloxane) is presented in Figure S9. Characteristic absorption peaks were evidenced in the region 1000–1130 cm^{-1}, which can be attributed to Si-O-Si stretching. Absorption peaks associated with $(CH_3)_2SiO$ groups were evidenced at 802 and 1258 cm^{-1}. The peaks at 2962 $cm^{-1}$are associated with methyl CH_3 stretching.[33]

Both types of MPET films were laminated with the glass fabrics using hand layup technique and all the laminates were found to exhibit a grammage of 490 ± 5 gsm, with a thickness 0.4 mm ± 0.05. The final configuration of the laminate prepared using dual-metallized PET is presented in Figure 7.

3.8 | Stiffness

In view of the final targeted application in the field of clothing, it is important to be cognizant about the flexibility of the laminate, which has been quantified in terms of the bending rigidity in both warp and weft directions. The bending length in both the directions and the overall flexural rigidity is presented in Table 3. The thermal shrinkage in all the laminates was found to be <1% at 260°C. In line with previous studies, the order of bending rigidity of the laminates prepared using woven fabrics followed the order satin < twill < plain.[36] Our studies indicate that of the three weave patterns, laminates prepared using satin fabrics are the most flexible and exhibit excellent potential to be used as outer layer of fire proximity clothing.

3.9 | Adhesion strength of laminates

In view of the flexible nature of both the substrates, the peel adhesion strength was determined as per ASTM D1876. The load-displacement plots associated with the peel test performed on representative laminates are presented as Figure 8. It can be observed that peel adhesion strength for the laminates prepared using dual-metallized PET films (irrespective of the OD) was found to be relatively higher (1.01 ± 0.03 N/mm), than that of single-metallized PET (0.63 ± 0.03 N/mm). This can be attributed to the stronger interaction of the poly(siloxane) adhesive with aluminium as compared to that with the polyester film. All the specimens fabricated displayed uniform peeling during the test. A photograph depicting the same is also included in the figure.

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Flexural rigidity of laminates prepared using metallized PET.

Sample designation	Bending length (cm)		Bending rigidity (mg.cm)		Overall flexural rigidity (mg cm)
	Warp	Weft	Warp	Weft	
ST-S-2.2	6.7	7.1	15105.1	18123.3	16545.8
ST-D-3.1	6.7	7.2	15241.1	18740.3	16900.1
ST-D-4.8	6.7	7.1	15038.1	18047.2	16474.1
TW-D-4.8	6.8	7.2	15860.7	19132.8	17420.1
PL-D-4.8	7.3	7.8	19772.4	23727.6	21659.9

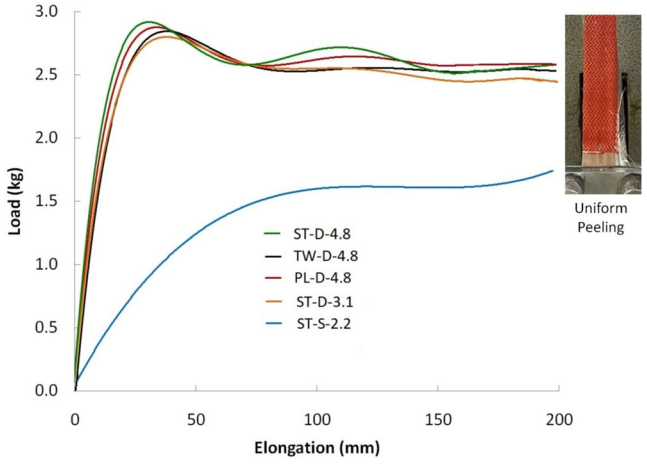


FIGURE 8 Load-displacement plots obtained from peel adhesion testing of different laminates.

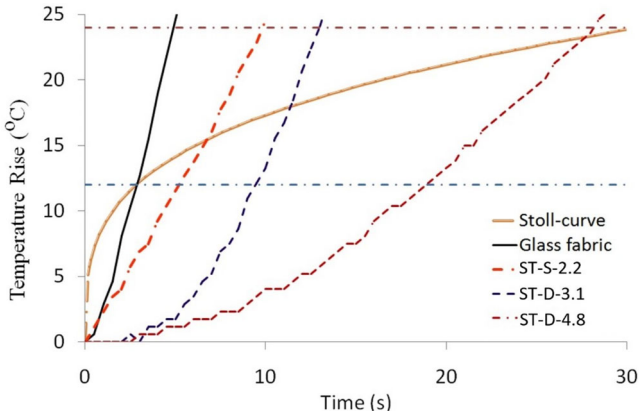


FIGURE 9 Effect of lamination and radiative exposure duration on the temperature rise.

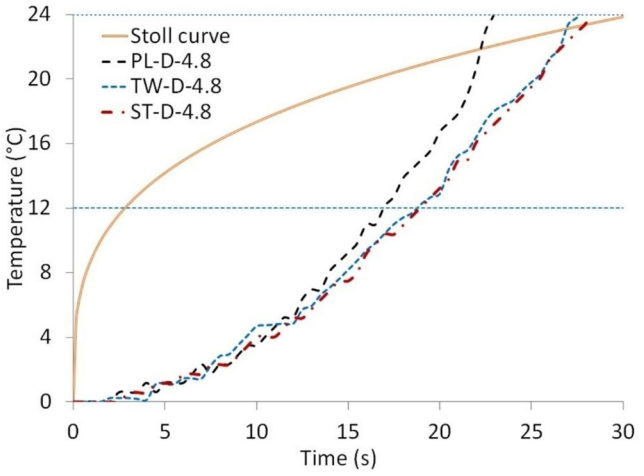


FIGURE 10 Effect of varying the weave pattern of the underlying glass substrate on the radiant protection level in terms of temperature rise with exposure duration (84 kW/m^2).

3.10 | Flexing under wet conditions

Fire proximity suits are prone to folding and puncturing; therefore, it is important to understand the effect of flexing on the properties of the laminate.[37] Amongst the several methods that have been used to simulate the mechanical stresses that laminates encounter during usage, the most widely used is Flex Testing, which is designed to evaluate the performance of laminates upon being subjected to reproducible mechanical stresses. The advantage of this stress simulation test is to reduce the cost and time associated with practical field tests. Flexing leads to the initiation of pinholes and stress cracks that could subsequently lead to a loss in barrier properties.[10] As observed during the T-peel tests, the laminates prepared using single-sided metallized PET films underwent significantly more damage during the wet flexing test as well. The photographs of the laminate prepared using both single- and dual-metallized PET after wet flex test is presented in Figure S10. It can be seen that the extent of damage was even more pronounced than in the dry state. The laminates prepared using single-sided metallized PET underwent delamination after ~500 cycles of flexing, while the dual-metallized PET layer adhered strongly after 1000 cycles even under wet conditions.

3.11 | Vertical flammability

One of the most important requirement of any material to find utility in fire protective clothing is its fire retardancy, which was established through vertical flammability testing. A photograph of a

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

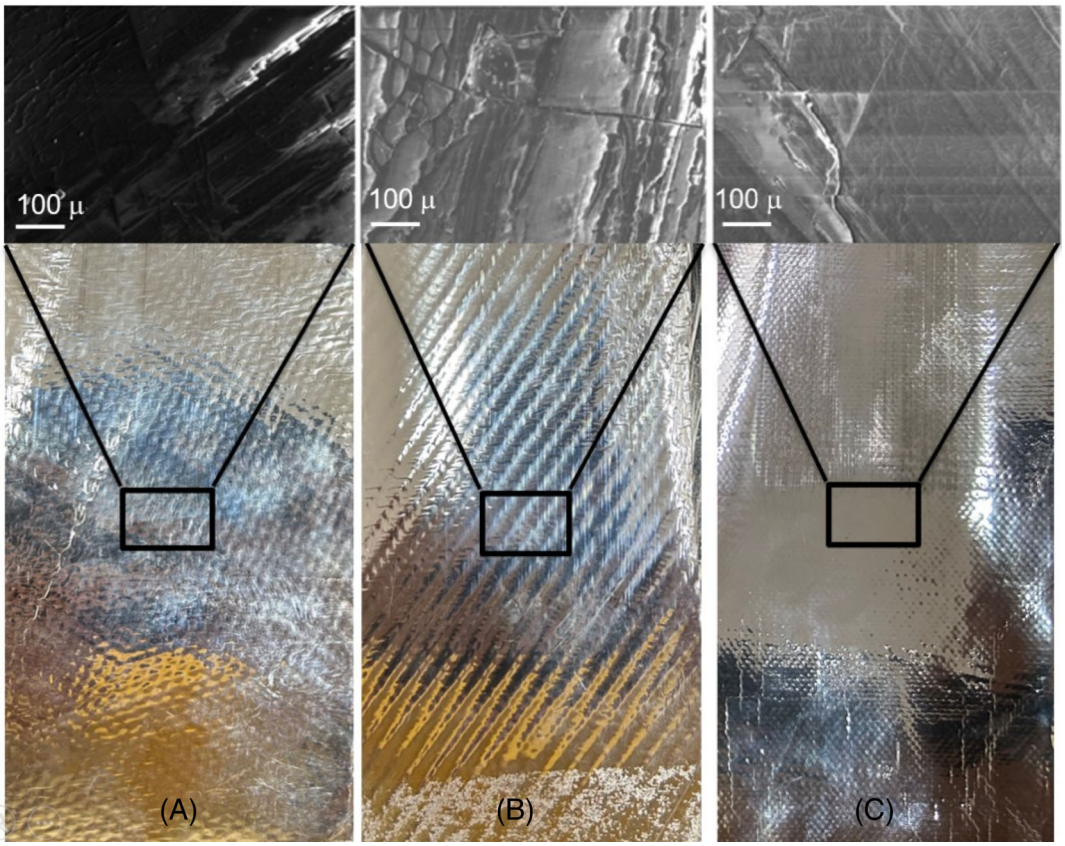


FIGURE 11 Photographs and SEM images of laminates post-radiative exposure based on (A) plain, (B) twill and (C) satin woven glass fabric.

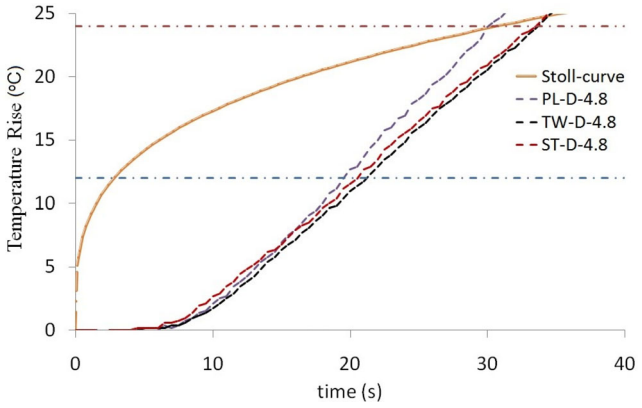


FIGURE 12 Response of ensemble in terms of temperature rise upon exposure to heat flux.

representative laminate post-testing is presented in Figure S11. It is to be noted that the MPET film when subjected to a vertical flame was totally engulfed in fire. However, post lamination, the laminates do not exhibit any after-flame or afterglow and there is no reportable char length. As per EN 1486,[11] for the outermost layer of the fire proximity clothing, the after-flame, after-glow and the char length requirement is ≤2.0 s, ≤2.0 s and ≤10 cm, respectively. In addition, melting or dripping of the sample is also not acceptable. Our studies reveal that all the laminates prepared meet the standard requirements.

3.12 | Radiant protection

The radiant protection is undoubtedly the most important criteria to be met by the laminate for possible use in the context of fire proximity clothing. For all practical purposes, the laminate needs to retain its reflectivity even after being subjected to a fixed number of abrasion cycles as mentioned in the standard. However, in the case of single side metallized film, this wearing process leads to a significant loss of the coating, which allows the radiative heat flux to be transferred. The effect of abrasion on these films was captured through digital imaging, the results of which are presented in Figure S12. It can be seen that significant erosion occurs in single-metallized films, which could be reduced substantially by dual metallization.

The effect of dual side metallization and increasing OD on the performance of the laminate upon exposure to fixed radiative heat flux (84 ± 2 kW/m^2) in terms of temperature rise as a function of increasing time duration is presented in Figure 9. It can be seen that the satin glass fabric is a rather poor shield against radiative heat, and it intersects the Stoll curve at 2.5 s. When laminated with a single-metallized PET film, the radiant performance of abraded film improved

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

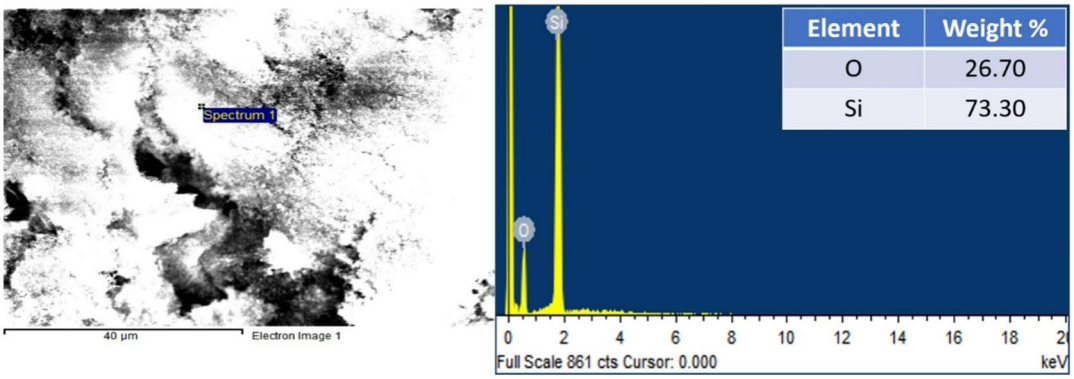


FIGURE 13 Elemental composition of the char formed as a result of convective exposure.

to 7.5 s; however, the requirement is not met. With dual metallization, the time to second degree burn increases further (t_{2nd} = 11 s), but only films with OD 4.8 were able to meet the requirements mandated by the standard (t_{2nd} > 20s).

3.13 | Effect of weave pattern on radiative exposure

Post confirming the ability of the high OD dual-metallized PET-glass laminate in meeting the standard requirements, it was considered prudent to investigate the effect of the weave pattern of the underlying substrate on the level of radiant protection offered. Previous studies have revealed that the surface roughness of the base fabric plays a crucial role in defining the level of protection offered against radiant heat. In this context, amongst the different aramid fabric structures studied, laminated non-woven fabrics were found to exhibit the highest RPP; however, they were unable to meet the requirement of mechanical strength mandated in the standards.[7] We hypothesized that smoother surfaces of satin and twill woven glass fabrics should offer higher level of radiant protection and the results of our studies are presented in Figure 10. Interestingly, both twill and satin woven fabrics were found to offer slightly higher protection against radiant heat flux as compared to plain weave configuration. Similar results have been previously obtained by Talukdar et al.[38]

It is to be noted that no visual damage was observed after 30 s of exposure in any of the laminate; however, the damage was evident in the SEM images (Figure 11). The SEM image clearly revealed the smooth surface of the laminated satin fabric.

3.14 | Thermal protective performance (TPP)

The TPP rating gives a quantitative estimate of the protection offered to the wearer upon exposure to heat flux.[39] The test was performed on the entire assembly of layers, which are used to fabricate the proximity suit. The photograph of the layers post-TPP testing is presented in Figure S13. Although the metallized PET reflective was completely destroyed due to exposure to convective heat, the underlying glass substrate remained intact, and so were the underlying substrate. It is interesting to note that upon exposure to radiant heat of similar heat flux for the same duration, the dual-metallized layer does not undergo any kind of damage which is visible to the naked eye; however, convective exposure leads to its complete destruction. It can be concluded that the role of the thermal barrier layer is more important in this type of exposure. The response in terms of temperature rise as a function of time is presented in Figure 12. It can be seen that irrespective of the weave pattern, all the ensembles exhibited tolerance time to second degree burn > 17.5 s (TPP = 35 cal/cm^2 s), which is a requirement of EN 1486. As observed in the case of radiant exposure, the laminate based on twill and satin woven fabric offer higher protection (tolerance time = 31 s, TPP= 62 cal/cm^2 s) as compared to plain woven fabric (tolerance time = 29 s, TPP = 58 cal/cm^2 s).

The SEM image and EDAX analysis of the char formed due to convective exposure is presented in Figure 13. As expected, the char formed comprised primarily of silicon and oxygen. The fluffy nature of the char could be considered responsible for preventing the transfer of heat to the underlying layers. Since laminate fabric prepared using satin fabric was observed to be more flexible than the twill and plain woven analogues, the former is considered to be the most suitable for application as the outermost layer in fire proximity suit.

4 | CONCLUSIONS

In the present work, the effect of dual side metallization and increasing OD of metallized PET films on its performance as the outermost reflective layer of fire proximity clothing was established. For any laminate to qualify the requirements of the outermost layer of fire proximity clothing, it should exhibit high level of reflectance as well as high degree of adhesive strength in addition to being flexible in both dry and wet conditions. Our studies reveal that a substantial fraction of the metal coating erodes during the abrasion cycle, which renders the laminate prepared using SMPET unsuitable for application as proximity outer layer. The abraded single-sided PET laminated on glass fabric (OD 2.1) is able to grant a

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

radiant protection from second degree burns for a period of ~7.5 s ($t_{2nd\ degree}$) upon exposure to radiant heat flux of 84 kW/m^2. In the dual-metallized analogue (OD 3.1), the metal coating present on the opposite face is able to effectively block the transfer of radiant heat, thereby increasing the protection level to ~11 s ($t_{2nd\ degree}$). By increasing the OD further to 4.8, the protection level could be increased to ~27 s, which is sufficient to meet the standard requirements. Furthermore, laminates prepared using single-sided metallized PET underwent delamination during the wet flexing tests and therefore were not suitable for fire proximity clothing applications. The weave pattern of underlying glass fabric was found to have a pronounced effect on the flexibility as confirmed by stiffness testing. Satin woven substrate was found to exhibit the lowest flexural rigidity followed by twill and plain weave substrates. Our studies lead us to conclude that for all practical purposes, dual-metallized PET film of high OD (4.8) laminated with both satin and twill woven glass fabric is the most suitable candidate for fire proximity clothing applications.

ACKNOWLEDGEMENTS

The authors deeply acknowledge to Dr. Ujjal K Gautam, IISER-Mohali for reflectance study, SRF polymers, Indore for preparation of metallized PET films and Montex Glass Fibre Industries Ltd for providing glass fabrics of different weaves.

CONFLICT OF INTEREST STATEMENT

All the authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Prasun Kumar Roy https://orcid.org/0000-0001-8459-1861

REFERENCES

1. Nayak R, Houshyar S, Padhye R. Recent trends and future scope in the protection and comfort of fire-fighters' personal protective clothing. *Fire Sci Rev*. 2014;3(1):4.
2. Xanthopoulos G, Delogu GM, Leone V, Correia FJM, Magalhães CG. Firefighting approaches and extreme wildfires. In: Tedim F, Leone V, McGee TK, eds. *Extreme Wildfire Events and Disasters: Root Causes and New Management Strategies*. Elsevier; 2019.
3. Omidvar B, Mohamadzadeh B. Fire event in oil, gas, and petrochemical industries. In: Rahimpour MR, Omidva B, Makarem MA, eds. *Crises in Oil, Gas and Petrochemical Industries: Disasters and Environmental Challenges*; 2023.
4. Abbott NJ, Schulman S. Protection from fire: nonflammable clothing — a review. *Fire Technol*. 1976;12:204-218.
5. Frydrych I, Cichocka A, Gilewicz P, Dominiak J. Comparative analysis of the thermal insulation of traditional and newly designed protective clothing for foundry workers. *Polymers (Basel)*. 2016;8(10):348.
6. Cai G, Wang H, Luo Z, Wang X. Study on thermal insulation properties of aluminised aramid fabrics. *Fibres Text East Eur*. 2015;23(4):52-56.
7. Jin L, Park PK, Hong KA, Yoon KJ. Effect of aluminized fabrics on radiant protective performance of fire proximity suit materials. *Ann Occup Hyg*. 2015;59(2):243-252.
8. Liu G, Liu Y, Zhao X. A study of the thermal protective performance of the outer fabric material for fire proximity suits. *J Text Inst*. 2018;109(7):851-857.
9. Kwon DJ, Kwon IJ, Milam-Guerrero JA, Yang SB, Yeum JH, Choi HH. Aramid nanofiber-reinforced multilayer electromagnetic-interference (EMI) shielding composites with high interfacial durability. *Maternite*. 2022;215:110452.
10. Ge C, Verma SS, Burruto J, Ribalco N, Ong J, Sudhahar K. Effects of flexing, optical density, and lamination on barrier and mechanical properties of metallized films and aluminum foil centered laminates prepared with polyethylene terephthalate and linear low density polyethylene. *J Plast Film Sheet*. 2021;37(2):205-225.
11. EN 1486:2007. *Protective Clothing for Fire-Fighters - Test Methods and Requirements for Reflective Clothing for Specialised Fire-Fighting*. 2007.
12. NFPA 1971. *Standard on Protective Ensembles for Structural Fire Fighting and Proximity Fire Fighting*. 2018.
13. Sharma P, Yadav A, Meena M, Roy P, Nebhani L. Epoxy-based reflective adhesive for lamination of single-sided metallized film over glass fabric destined for its use as reflective layer in fire proximity clothing. *Polym Degrad Stab*. 2023;213:110365.
14. ASTM E96. *Standard Test Methods for Water Vapor Transmission of Materials*. 2005.
15. ASTM F2894-21. *Standard Test Method for Evaluation of Materials, Protective Clothing, and Equipment for Heat Resistance Using a Hot Air Circulating Oven*. 2021.
16. ASTM D 1388. *Standard Test Method for Stiffness of Fabrics 19428–2959*. 2018.
17. ASTM D 1876-01. *Standard Test Method for Peel Resistance of Adhesives (T-Peel Test) 19428–2959*. 2001.
18. ASTM D 6413. *Standard Test Method for Flame Resistance of Textiles (Vertical Test) 19428–2959*. 2013.
19. ASTM D 410. *Standard Test Method for Thermal Protective Performance of Materials for Clothing by Open-Flame Method 19428–2959*. 1987.
20. Culbertson E. Metal Adhesion to PET Film. Association of Industrial Metallizers, Coaters and Laminators - Fall Technical Conference and 20th International Vacuum Web Coating Conference 2006 2006.
21. Niazmand R, Yeganehzad S, Niazmand A. Application of laminated and metalized films to prolong the shelf life of dried barberries. *J Stored Prod Res*. 2021;92:101809.
22. De Puydt Y, Bertrand P, Lutgen P. Study of the Al/PET interface in relation with adhesion. *Surf Interface Anal*. 1988;12(9):486-490.
23. Bou J, martin J, Le Mogne T, Vovelle L. XPS study of the interface between thermally evaporated aluminium and polyethylene terephthalate: evidence for oxycarbide species. *Appl Surf Sci*. 1991;47:219-231.
24. Sandrin L, Sacher E. X ray photelectron spectroscopy studies of the evaporated aluminium /corona treated polyethylene terephthalate interface. *Appl Surf Sci*. 1998;135(1):339-349.
25. Rancourt JD, Hollenhead JB, Taylor LT. Chemistry of the Interface between aluminum and polyester films. *J Adhes Dent*. 1993;40:40.
26. Travaly Y, Bertrand P. Static SIMS investigation of metal/polymer interfaces. *Surf Interface Anal*. 1995;23:328-334.
27. Feldman CL, Mayer JW. *Fundamentals of Surface and Thin Film Analysis*. Elsevier; 1984.
28. Gupta S, Dixit M, Sharma K, Saxena NS. Mechanical study of metallized polyethylene terephthalate (PET) films. *Surf Coat Technol*. 2009;204(5):661-666.
29. Quintiere J. Radiative characteristics of fire fighters' coat fabrics. *Fire Technol*. 1974;10(2):153-161.

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

30. Zhu FL, Feng QQ. Preparation, thermal properties and permeabilities of aluminum-coated fabrics destined for thermal radiation protective clothing. *Fire Mater*. 2020;44(6):844-853.
31. Hao L, Yu W. Comparison of thermal protective performance of aluminized fabrics of basalt fiber and glass fiber. *Fire Mater*. 2011;35(8):553-560.
32. Meena M, Roy P, Jacob J. Anti-drip flame retardant poly(urethane-imide) thin films for high-temperature applications. *J Appl Polym Sci*. 2023;140(33):54297.
33. Kim EY, Kong JS, An SK, Do KH. Surface modification of polymers and improvement of the adhesion between evaporated copper metal film and a polymer. I. Chemical modification of PET. *J Adhes Sci Technol*. 2000;14(9):1119-1130.
34. Kerketta A, Meena M, Dwivedi S, Yadav R, Roy P. Tuning the properties of breathable polyurethanes through chemical crosslinking. *J Appl Polym Sci*. 2023;20:54263.
35. Lin SB, Durfee LD, Ekeland RA, McVie J, Schalau GK. Recent advances in silicone pressure-sensitive adhesives. *J Adhes Sci Technol*. 2007;21(7):605-623.
36. Punj SK, Mukhopadhyay A, Pattanayak A. Flexural rigidity of different weave structures. *Text Asia*. 2002;33(6):33-36.
37. Lange J, Büsing B, Hertlein J, Hediger S. Water vapour transport through large defects in flexible packaging: modeling, gravimetric measurement and magnetic resonance imaging. *Packag Technol Sci*. 2000;13(4):139-147.
38. Talukdar P, Das A, Alagirusamy R. Effect of structural parameters on thermal protective performance and comfort characteristic of fabrics. *J Text Inst*. 2017;108(8):1430-1441.
39. Zhang X, Tian M, Huang W, Su Y, Li J. Effects of under-clothing airflow on thermal protective performance and thermal aging behaviors of flame-retardant fabric. *Fire Mater*. 2023;47(8):1053-1063.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Dwivedi S, Srivastava R, Kumar Roy P. Exploring the potential of dual-metallized PET towards improving the efficiency of outermost reflective layer in fire proximity clothing. *Fire and Materials*. 2024;1-13. doi:10.1002/fam.3229

10991018, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/fam.3229 by Delhi Technological University, Wiley Online Library on [08/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 24 2024

Exploring the wear behaviour of Al6351-Gr composite reinforced with B_4C, TiC, and WC particles

Shahazad Ali ✉; Qasim Murtaza; Pallav Gupta

Check for updates

AIP Conf. Proc. 3111, 060001 (2024)
https://doi.org/10.1063/5.0221452

View Online
Export Citation

06 August 2024 04:50:07

Nanotechnology & Materials Science
Optics & Photonics
Impedance Analysis
Scanning Probe Microscopy
Sensors
Failure Analysis & Semiconductors



Unlock the Full Spectrum.
From DC to 8.5 GHz.
Your Application. Measured.
Find out more
Zurich Instruments

AIP Publishing

Exploring the Wear Behaviour of Al6351-Gr Composite Reinforced with B_4C, TiC, and WC Particles

Shahazad Ali[1, a)] Qasim Murtaza[1,b)] Pallav Gupta[2,c)]

[1]*Mechanical, Production and Industrial Engineering Department, Delhi Technological University, Delhi, 110042, India*

[2]*Department of Mechanical Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida 201313, India.*

Corresponding author: [a)]shahzadali0786@gmail.com, [b)]qasimmurtaza@dce.ac.in, [c)]pgupta7@amity.edu

Abstract. The wear behavior of Al6351-Gr composites reinforced with B_4C, WC, and TiC were investigated in this study. Wear resistance is a critical factor in determining the functionality and durability of materials in different industrial relevance. Hybrid composites were created using a liquid metallurgy technique. The samples were subjected to wear tests using a pin-on-disk apparatus under controlled experimental conditions. Wear performance, friction coefficient, and wear processes were evaluated and compared between the composite materials. According to the result, the addition of B4C, TiC, and WC to the Al6351-Gr composite significantly improved its wear resistance. The examination of the worn surfaces through scanning electron microscopy (SEM) unveiled a diverse range of wear mechanisms, encompassing both adhesive wear and abrasive wear. The reinforced composite samples exhibited reduced wear severity and extent compared to the unreinforced material, confirming the improvement in their wear resistance.

06 August 2024 04:50:07

INTRODUCTION

Composite materials are specifically designed materials comprising two or more constituent substances possessing distinct physical or chemical characteristics. These materials are designed to take advantage of the unique properties of each constituent material to achieve specific performance characteristics that cannot be obtained with any single material alone. The constituent materials of composites are typically classified as either matrix materials or reinforcement materials.

In composite materials, the matrix material acts as a continuous phase, enveloping and providing support to the reinforcement material, which, in turn, imparts strength and rigidity to the composite structure. The reinforcement material can manifest in various forms, such as fibers, particles, or flakes.

Composite materials possess the advantageous characteristic of combining multiple desirable properties, including but not limited to exceptional strength, lightweight nature, resistance to corrosion, and thermal stability. As a result, they find extensive utilization across various industries, such as aerospace, automotive, construction, and sporting goods, due to their versatility and broad range of applications.

Silicon carbide is an effective ceramic material that enhances the wear properties of hybrid MMC. When combined with Graphite, which possesses excellent self-lubricating properties, the addition of Graphite, along with other ceramics, further enhances the wear characteristics of the hybrid MMC. Moreover, the inclusion of Silicon carbide helps maintain the desired hardness properties of the hybrid MMC [1][2][3].

Bhowmik et al. conducted a study on the wear properties of Al7075 by incorporating various proportions of SiC through the stir-casting process. During the stirring process, XRD analysis revealed the presence of silicon carbide particles well dispersed within the aluminum matrix. The wear rate was observed to increase, while the coefficient of friction (COF) decreased with higher sliding speeds, primarily due to intense friction between the pin and disc. Examination of the worn surface of the pin indicated that the Al7075 reinforced with 9% SiC composite exhibited the least material removal [4]. Furthermore, it should be noted that not only the presence of SiC but also its particle size, significantly influences the wear characteristics[5]. Other studies focusing on SiC reinforcement have also demonstrated promising improvements in wear properties in aluminum alloys[6].

Naseem et al. conducted a study on the wear properties of Al6351 by incorporating a combination of Al_2O_3 and Carbon. The worn surface was analyzed using a scanning electron microscope, which revealed that the pure Al matrix exhibited deep grooves, higher debris accumulation, and a rough surface compared to the composites with a

AICTE & DSIR Sponsored International Conference on Smart and Innovative Development in Science, Engineering & Technology 2023
AIP Conf. Proc. 3111, 060001-1–060001-9; https://doi.org/10.1063/5.0221452
Published under an exclusive license by AIP Publishing. 978-0-7354-5008-0/$30.00

higher percentage of reinforcement. Interestingly, as the percentage of reinforcements increased, the wear rate of the hybrid composite decreased [7]. Adding SiC or TiO_2along with Al_2O_3 and graphite shows further enhancement in wear attributes of hybrid MMCs[8][9], [10][11].

Mohanavel et al. developed a composite by incorporating varying proportions of Si_3N_4 into Al6351 and investigated its wear characteristics. The inclusion of Si_3N_4 particles had a significant impact on the hardness of the original material, leading to a reduction in the material's wear rate. Furthermore, the strong bond between the Al6351 matrix and the Si_3N_4 ceramic reinforcement particles prevented the detachment of Si_3N_4 material during sliding. Consequently, the Al6351 composite reinforced with 3 wt.% Si_3N_4 exhibited the lowest wear rate compared to the unmodified alloy [12]. The inclusion of TiB_2 and ZrB_2 in aluminum-based alloys significantly enhances both the wear resistance and mechanical properties of the alloys. [13][14].

Numerous research studies have examined the wear characteristics of composites reinforced with graphene nanoplatelets (GNPs). These investigations have consistently shown that the inclusion of GNPs significantly enhances the wear endurance of the composites. This improvement can be attributed to the exceptional properties of graphene, including its high strength and low friction [15][16][17].

Rice husk ash and eggshells, which are considered waste materials, are commonly employed to improve the mechanical properties and wear characteristics of aluminum-based hybrid composites. The addition of these reinforcements not only yields beneficial results but also provides the advantage of cost-effectiveness [18][19].

This study aims to explore the effects of incorporating B_4C, TiC, and WC as reinforcements in an Al6351-Gr composite on its wear characteristics. While previous studies predominantly focused on investigating high proportions of reinforcement, this study specifically aims to explore the effects of a lower proportion of reinforcement.

RESEARCH OBJECTIVES

The primary goal of this research is to stir cast Al6351&Graphite composite reinforced with B4C, TiC, and WC. The specific goal of this research is:

1. To assess the wear functionality of the prepared hybrid MMC using a pin-on-disk wear testing machine.
2. To analyze the microphotograph of worn surfaces.
3. To propose potential applications for hybrid composite alloy matrix hybrid composites based on their performance characteristics.

06 August 2024 04:50:07

MATERIALS AND METHODS

Materials

To fabricate a hybrid metal matrix, aluminum ingots with a were utilized as the matrix metal. Graphite, B_4C, TiC, and WCwere used as reinforcement materials. Graphite is a lightweight and strong structure material with excellent thermal conductivity[20].

B_4C is a low-density ceramic with exceptional hardness and high wear resistance. Its high hardness makes it effective in reducing abrasive wear by acting as a protective barrier against contacting surfaces. B_4C has a low coefficient of friction, which can help reduce frictional wear between surfaces. It can form a self-lubricating film during wear, reducing the tendency for adhesive wear.

TiC is a ceramic material known for its exceptional hardness, strength, and resistance to wear. It provides high wear resistance due to its hardness, effectively reducing abrasive wear. TiC has a high melting point, making it suitable for applications involving high temperatures or frictional heat generation. It can improve the load-bearing capacity and endurance of composites by reinforcing the metal matrix.

WC is a hard compound composed of tungsten and carbon atoms, known for its extreme hardness and wear resistance. It possesses excellent resistance to abrasive wear, making it suitable for applications involving harsh conditions. WC particles can reinforce the composite matrix, improving its strength, toughness, and wear resistance.

Table 1 displays the physical properties of Graphite, B4C, WC, and TiC.

TABLE 1. Physical Characteristics of Reinforcement

Reinforcement	Weight density(gram/cc)	Particle size (micron)	Color

Graphite (Gr)	1.35	40-80	Gray black
Boron Carbide (B4C)	2.52	80-120	Grayblack
Tungsten Carbide (WC)	15.7	40-80	Gray
Titanium Carbide (TiC)	4.91	80-120	Silver gray

The terminology and composition of the fabricated hybrid MMCs are shown in Table 2.

Table 2. Sample Nomenclature of AlGr hybrid MMCs

S.N	Al6351 (wt. %)	Reinforcement (wt. %)	Sample Code
1	100.00	0 wt %	AlGr
2	98.0	1.00 wt % Gr + 1.00 wt % B4C	AlGrB
3	98.0	1.00 wt % Gr + 1.00 wt % TiC	AlGrT
4	98.0	1.00 wt % Gr + 1.00 wt % WC	AlGrW

Fabrication Technique

The fabrication of the Al6351 & Graphite composite reinforced with B4C, TiC, and WC using the stir casting process involves the following steps:

In the very beginning, the materials were prepared according to the composition shown in Table 2. AA6351 was cut into small pieces and according to the weight of Al6351, the reinforcement proportion was measured. Melting of the Al6351 aluminum alloy was carried out in a temperature-controlled muffle furnace at 930°C.Reinforcements as per proportions were heated in another horizontal muffle furnace to remove moisture content. After Approx 40 minutes, the temperature of the vertical furnace was increased to 960°C and the molten mixture was stirred for five minutes to get uniformity. Then after, gradually reinforcement particles are introduced into the molten matrix material. The Stirring of the mixture was carried out continuously using a mechanical stirrer to ensure uniform distribution of the particles throughout the molten matrix. A degassing agent was used to remove any trapped gases or air bubbles from the molten mixture.The stirred molten composite mixture was poured into the preheated mold and allowed molten composite to cool and solidify inside the mold. The solidification process was carried out in natural air cooling.Once the composite solidified and reached sufficient strength, it was removed from the mold carefully, and postprocessing finishing processes such as machining, grinding, and polishing were carried out to achieve the desired shape, dimensions, and surface finish. A similar process was repeated for fabricating other hybrid composites.

The stir-casting process allows for the incorporation of reinforcement particles into the matrix material, resulting in a homogeneous distribution of the particles within the composite. The process offers flexibility in controlling the composition, particle size, and volume fraction of the reinforcement materials. It is a widely used and cost-effective method for fabricating metal matrix composites.

EXPERIMENTATION

After solidification, samples were gone through various machining processes. Samples were prepared for mechanical and wear tests. Then after the samples go through the various tests.

Mechanical Testing

The fabricated samples underwent various tests to evaluate their mechanical properties. Tensile strength was measured using a PC2000 Kundle tensiometer, hardness was determined using a hardness testing machine, and the impact strength of the hybrid composite was assessed through a Charpy test. The specimen preparation followed the

06 August 2024 04:50:07

guidelines outlined in ASTM standards. The results obtained from the tensile test, hardness test, and impact test are consolidated in the table below.

TABLE 3. Mechanical characterization of fabricated samples

Sample	Engg. UTS (MPa)	Hardness (HRC)	Toughness (Joules)
AlGr	100	69	26
AlGrB	134	82	32
AlGrW	154	76	25
AlGrT	206	85	33

Sample AlGrT shows the best mechanical characteristics among all the fabricated samples due to uniform proper mixing and lower defects.

WEAR TEST

The wear behavior of the fabricated hybrid MMCs was evaluated using a pin-on-disk wear testing machine. The wear testing of the Al6351 & Graphite composite reinforced with B4C, TiC, and WC involves the use of an experimental setup designed to simulate the wear conditions and measure the wear behavior of the composite material. The wear testing was performed using a pin-on-disk apparatus, which consists of a rotating disk and a stationary pin. The composite sample, in the form of a pin, was brought into contact with the rotating disk under controlled conditions. Below explained parameters and test conditions were considered for wear experiments:

- Applied Load: The load or force applied to the composite material during wear testing is an important parameter. Load in the range of 10 to 40 N varied with equal intervals of 10 to simulate different levels of contact pressure and determine the load-bearing capacity and wear resistance of the composite.
- Sliding Speed: The sliding speed or rotational speed of the test setup is another critical parameter. It represents the relative motion between the composite sample and the counterface material. Sliding speeds can range from low-speed conditions to higher speeds, depending on the intended application.The tests were performed with a consistent track diameter of 65 mm on the disc. The disc's speed varied at 295 RPM during the experiments to get a linear relative velocity of 1m/s.
- Sliding Distance: The sliding distance represents the cumulative distance traveled by the sliding contact. It is an essential parameter for assessing the wear resistance and durability of the composite. The workpieces were subjected to a sliding distance of 1000 m to achieve a noticeable weight loss in the specimens.
- Counterface Material: The choice of counterface material, which encounters the composite sample, is crucial. It should represent the intended application or the typical wear conditions that the composite will encounter. Based on the recommendation of ASTM, EN31 steel was selected due to its hardness and compatibility with the composite material.

Experimental Procedure

To conduct a wear test on the fabricated hybrid MMCs, the following steps were implemented. Firstly, the experimental setup was prepared by meticulously cleaning the specimen pin and eliminating any burns on the circumference using fine-grade emery paper. The disc was also thoroughly cleansed with a solvent and allowed to dry to assess wear on the disc surface. The weight of the specimen was then accurately measured using a highly precise weighing machine. Subsequently, the sample pin was securely clamped to the loading lever tip using hardened jaws of appropriate size. To establish a wear radius of 65 mm, the sliding plate was adjusted by unscrewing and shifting it to the desired position, aligning it with the graduated scale, and then securing it in place with the clamp screw. The machine operated at a speed of 295 RPM, and the process continued for the remaining duration while monitoring for any fluctuations.

sliding velocity and sliding distance were calculated by using below equation:

$$\text{Sliding speed in m/sec} = \frac{\pi DN}{60000}$$

06 August 2024 04:50:07

$$\text{Sliding distance in meters} = \frac{\pi DNT}{60000}$$

Where D is the diameter of the wear track in mm, N is the disc speed in rpm and T is the test duration in seconds.

After setting the rotation of the machine, the normal load was applied slowly withoutshaking. The experiment was conducted for all the specimens, and the data was observed.During and after experimentation weight loss and coefficient of friction data were collected to analyze the wear behavior of the hybrid MMCs.

Wear Experimental Result and Analysis

The specimens were subjected to a wear testing machine to assess their performance. Initially, a low load of 10N was applied to each specimen. After traveling a distance of 1000 m, the weight losses and coefficients of friction were measured. Subsequently, the load was incrementally increased to 20N, 30N, and 40N, and the corresponding weight losses and coefficients of friction were recorded and presented in the table. The wear rate data is summarized in Table 4, and its graphical representation is depicted in Fig. 1.

TABLE 4. Wear rate of samples (mm³ /m)

Load (N)	Al6351	AlBG	AlWG	AlTG
10	0.000752	0.000512	0.000612	0.000498
20	0.002564	0.0018688	0.002098	0.0013031
30	0.006783	0.004806	0.005589	0.0042179
40	0.007621	0.00632	0.006821	0.00592

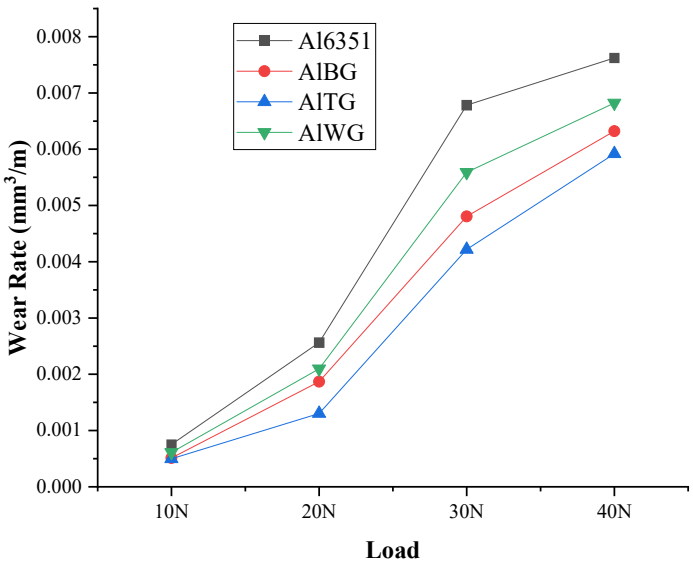


FIGURE 1. Wear Rate of Fabricated Samples

Table 5 displays the coefficient of friction along with the respective load, and Fig. 2 depicts its graphical representation.

TABLE 5. Coefficient of friction Vs load data

Load (N)	Al6351	AlBG	AlWG	AlTG
10	0.37	0.3321	0.3521	0.3212

06 August 2024 04:50:07

20	0.39	0.355206	0.3789	0.341638
30	0.412	0.37577	0.3932	0.366018
40	0.4321	0.387854	0.4121	0.374286

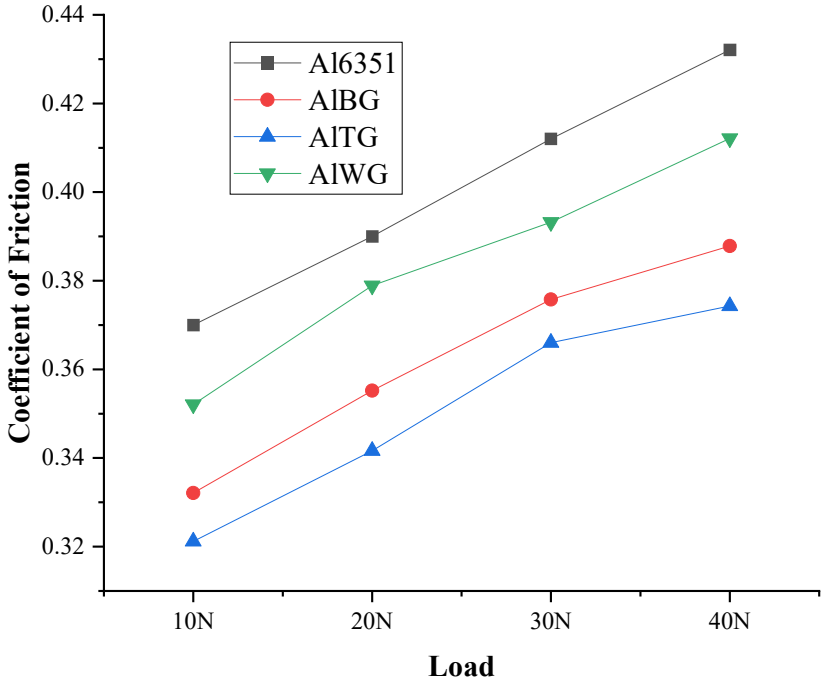


FIGURE 2. Variation of friction with load

In accordance with Archard's equation, the wear resistance of a material is inversely related to its hardness. By increasing the hardness of composite material, its ability to withstand sliding wear is improved. The addition of reinforcement particles, which are inherently hard, reduces the effective contact area between Al6351 and the disc. This redistribution of contact area, where some of it is occupied by the ceramics particles, enhances the wear resistance of the composite. Furthermore, the strong bond between the matrix and the ceramics particles prevents their separation during sliding, further enhancing the wear resistance of the hybrid MMC compared to the monolithic alloy. Figure 1and Fig. 2 illustrate the impact of load on the wear rate of the unreinforced Al alloy and Hybrid MMCs. Notably, the Al6351 matrix reinforced with 1 wt% TiC and 1 wt.% consistently exhibits a lower wear rate under all load conditions.

Worn Surface Analysis

Following the experimentation, the worn surfaces were studied using a high-resolution microscope (SEM). These detailed images serve as a valuable tool for identifying wear mechanisms, observing wear tracks, and analyzing wear debris.Fig. 3 showcases the impact of reinforcement content on the worn microstructure of the composites under a normal load of 40 N. The worn surface microstructure of the Al6351 alloy displays noticeable plastic deformation.Heat is generated between the specimen to be tested and the countersurface due to their relative motion and friction between them. The wear mechanism is predominantly adhesive. In Fig. 3(b, c, d), the worn surface micrographs of the composites display narrower grooves and a smaller number of pits. The presence of reinforcement particles reduces the depth of the grooves and minimizes plastic deformation along the groove edges. The wear mechanism in this case is primarily abrasive.

06 August 2024 04:50:07

In Fig. 3(a), the observed microstructure reveals several grooves and scratches that are oriented along the sliding direction. This observation provides significant insights into the predominant wear characteristics, which are primarily attributed to adhesion and delamination.

The presence of multiple grooves suggests the occurrence of adhesive wear, where the contact surfaces experience strong forces leading to material transfer and subsequent formation of grooves. This phenomenon is often associated with the intimate contact and bonding between the mating surfaces, resulting in material removal and the formation of wear debris.

Additionally, the presence of scratches further indicates the involvement of delamination wear. Delamination occurs when layers of material separate or peel off from the surface due to shear forces during sliding. These scratches can be considered evidence of localized material detachment and failure, which contributes to the overall wear process.

Figures 3(b), 3(c), and 3(d) showcase the impact of adding reinforcement particles on the worn surface characteristics. These figures reveal that adding the reinforcement particle restricts plastic deformation and minimizes the formation of grooves on the worn surface. Notably, the movement of metal on the worn surfaces of the hybrid Metal Matrix Composites (MMCs) is substantially reduced compared to the unreinforced Al6351 alloy.

Among the different hybrid MMC compositions investigated, the hybrid MMC with a 1wt% reinforcement particle content exhibits the least number of debris and a smoother worn surface when compared to the other hybrid MMC variants. This indicates that a 1wt% reinforcement particle content effectively mitigates wear and results in improved surface quality.

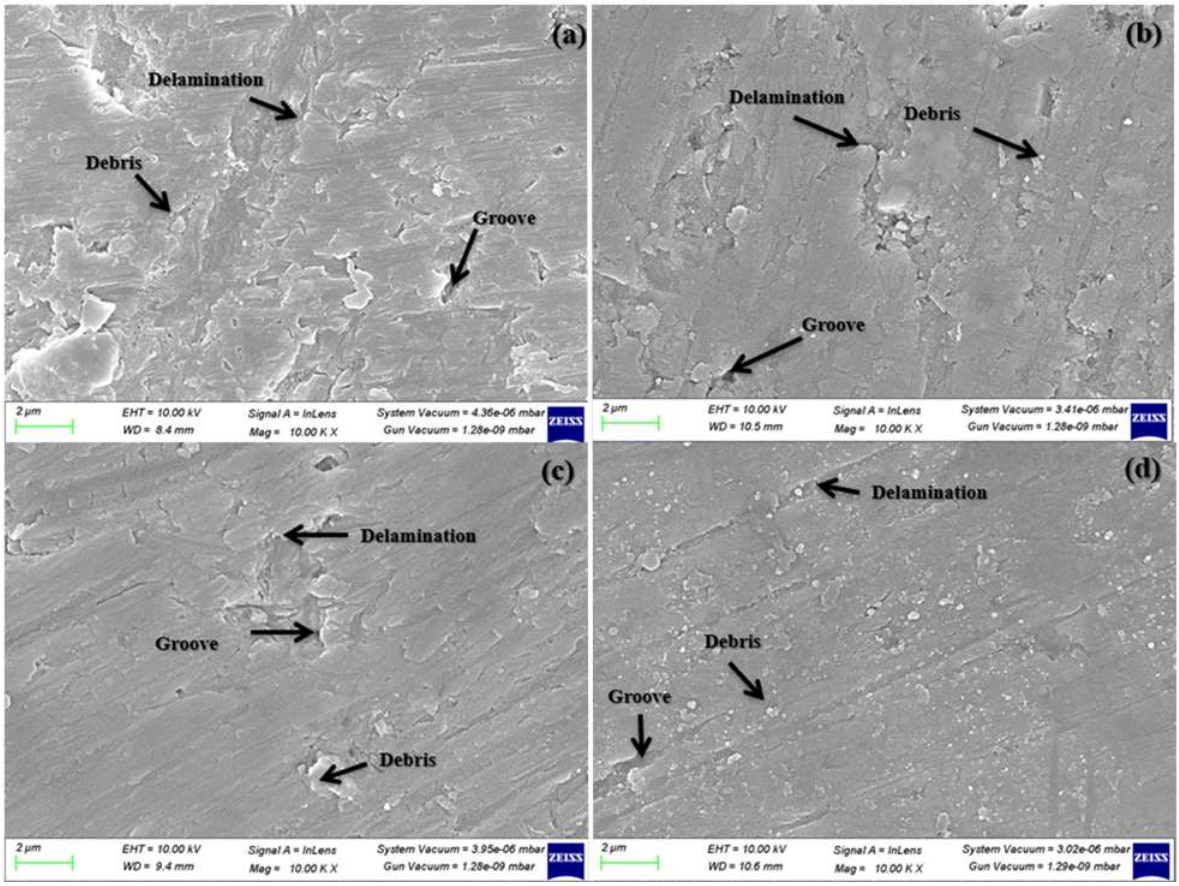


FIGURE 3.SEM of Worn Surfaces

Based on the improved corrosion and wear resistance of the Gr-TiC-Al6351 alloy matrix hybrid composites, potential applications include marine and aerospace industries, where corrosion and wear resistance are critical. Composites can also be used in automotive and mechanical engineering applications, where improved performance is desired.

06 August 2024 04:50:07

CONCLUSION

The present study explored the wear performance of Al6351-Gr reinforced with B4C, WC, and TiC. The hybrid MMCs were fabricated using a stir-casting process. The wear test revealed a consistent improvement in the wear performance of the fabricated HMMCs in comparison tothe base Al6351 alloy. The highest wear rate was observed in Al6351 as 0.007621 mm^3/m and the lowest in AlGrT (1 wt.% Gr& 1 wt.% TiC) as 0.00592 mm^3/m. The coefficient of friction shows an increasing pattern with increasing load. The lowest friction was observed for sample AlGrT at 40 N as 0.374286.

In microphotograph, Al6351 shows a large number of debris and plowing. The addition of reinforcement shows a reduction of debris. Sample AlGrT shows the least number of debris and plowing.

REFERENCES

1. S. H. Dhoria, V. Durga Prasada Rao, and K. Venkata Subbaiah, Mechanical and wear behaviour of 6351 Al/Gr/SiC composites fabricated by squeeze casting (Materials Today: Proceedings, vol. 18, 2019), pp. 2107–2113.
2. V. Mohanavel, S. Suresh Kumar, P. Ganeshan, R. Vishnukumar, and P. Gurusamy, Physical, mechanical and wear performance of graphite (Gr) particles incorporated aluminium matrix composites – An overview (AIP Conference Proceedings (Vol. 2283, No. 1). AIP Publishing, vol. 2283, no. 1, 2020).
3. K. Vijaya Bhaskar, S. Sundarrajan, B. Subba Rao, and K. Ravindra, Effect of reinforcement and wear parameters on dry sliding wear of aluminum composites-A review (Materials Today: Proceedings, vol. 5, no. 2,2018) pp. 5891–5900.
4. A. Bhowmik, D. Chakraborty, D. Dey, and A. Biswas, Investigation on wear behaviour of Al7075-SiC metal matrix composites prepared by stir casting (Materials Today: Proceedings, vol. 26, 2019), pp. 2992–2995.
5. Q. Zhao, B. Ju,K. Zhao, J. Zhan, M. Liu, N. Zhang, J. Qian,Z. Xiu, P. Kang, and W. Yang, Effect of Reinforcement Size on Mechanical Behavior of SiC-Nanowires-Reinforced 6061Al Composites (Materials, vol. 15, no. 23, 2022).
6. V. Mohanavel, K. Rajan, S. Suresh Kumar, A. Chockalingam, A. Roy, and T. Adithiyaa, Mechanical and tribological characterization of stir-cast Al-SiCpcomposites (Materials Today: Proceedings, vol. 5, no. 1,2018) pp. 1740–1746.
7. N. Ahamad, A. Mohammad, and P. Gupta, Wear characteristics of Al matrix reinforced with Al2O3-carbon hybrid metal matrix composites, (Materials Today: Proceedings, vol. 38, 2021) pp. 63–68.
8. C. S. Kalra, V. Kumar, and A. Manna, The wear behavior of Al/(Al2O3 + SiC + C) hybrid composites fabricated stir casting assisted squeeze (Particulate Science and Technology, vol. 37, no. 3, 2019) pp. 303–313.
9. V. Mohanavel, K. Rajan, S. Sureshkumar, and P. Ganeshan, Manufacturing and mechanical characterization of dual particle reinforced hybrid composites, (Proceedings of 2017 International Conference on Intelligent Computing and Control, I2C2 2017, vol. 2018) pp. 1–5.
10. V. Mohanavel, K. Rajan, M. Ravichandran, S. Suresh Kumar, M. Balamurugan, and C. Jayasekar, Physical and tribological behaviour of dual particles reinforced metal matrix composites (Lecture Notes in Mechanical Engineering. Springer, Singapore, 2019).
11. N. Ahamad, A. Mohammad, K. K. Sadasivuni, and P. Gupta, Wear, optimization and surface analysis of Al-Al2O3-TiO2 hybrid metal matrix composites (Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, vol. 235, no. 1, 2021) pp. 93–102.
12. V. Mohanavel, K. S. A. Ali, S. Prasath, T. Sathish, and M. Ravichandran, Microstructural and tribological characteristics of AA6351/Si3N4 composites manufactured by stir casting (Journal of Materials Research and Technology, vol. 9, no. 6, 2020) pp. 14662–14672.
13. S. Suresh and N. S. V. Moorthi, Process development in stir casting and investigation on microstructures and wear behavior of TiB2 on A16061 MMC (Procedia Engineering, Elsevier Ltd, 2013), pp. 1183–1190.
14. G. Naveen Kumar, R. Narayanasamy, S. Natarajan, S. P. Kumaresh Babu, K. Sivaprasad, and S. Sivasankaran, Dry sliding wear behaviour of AA 6351-ZrB2 in situ composite at room temperature (Materials & Design, vol. 31, no. 3, 2010) pp. 1526–1532.
15. S. Zhang, Z. Chen, P. Wei, W. Liu, Y. Zou, Y. Lei, Y. Lei, S. Yao , S. Zhang, B. Lu, and L. Zhang, Wear properties of graphene/zirconia biphase nano-reinforced aluminium matrix composites prepared by SLM (Materials Today Communication, vol. 30,2022).

06 August 2024 04:50:07

16. R. R. Raj, J. Yoganandh, M. S. S. Saravanan, and S. S. Kumar, Effect of graphene addition on the mechanical characteristics of AA7075 aluminium nanocomposites (Carbon Letters, vol. 31, no. 1, 2021) pp. 125–136.
17. M. Naik H R, M. L H, V. Malik, M. Patel GC, K. K. Saxena, and A. Lakshmikanthan, Effect of microstructure, mechanical and wear on Al-CNTs/graphene hybrid MMC'S (Advances in Materials and Processing Technologies, 8:sup2, 2022) pp- 366-379.
18. G. Arora and S. Sharma, A Comparative Study of AA6351 Mono-Composites Reinforced with Synthetic and Agro Waste Reinforcement (International Journal of Precision Engineering and Manufacturing, vol. 19, no. 4, 2018) pp. 631–638.
19. Z. Sydow, M. Sydow, Ł. Wojciechowski, and K. Bieńczak, Tribological performance of composites reinforced with the agricultural, industrial and post-consumer wastes: A review, (Materials, vol. 14, no. 8. MDPI AG, 2021).
20. S. Ali, Q. Murtaza, and P. Gupta,Synergetic Effect of Gr-B4C Reinforcement on the Structural and Mechanical Properties of AA6351 Hybrid Metal Matrix Composites (ECS Journal of Solid State Science and Technology, vol. 12, no. 6, 2023).

06 August 2024 04:50:07

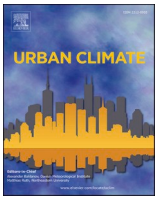
Urban Climate 56 (2024) 102066



Contents lists available at ScienceDirect

Urban Climate

journal homepage: www.elsevier.com/locate/uclim





Geographical variability of ultrafine particle concentrations in urban and background regions in India

Vignesh Mohan [a], Vijay Kumar Soni [b], Rajeev Kumar Mishra [a,*]

[a] Department of Environmental Engineering, Delhi Technological University, Delhi, 110042, India
[b] India Meteorological Department, Ministry of Earth Sciences, New Delhi, 110003, India

ARTICLE INFO

Keywords:
Air pollution
Particle number concentration in India
Regional background
Urban ultrafine particles

ABSTRACT

The study measured and analysed the particle number concentration of particles ranging from 10 to 800 nm in urban and background sites. The aim of the study is to estimate the role of anthropogenic sources in urban regions in terms of number concentration. The study finds that the particle number concentration in an urban site is ~8 times higher than in the background sites. In the urban site, the Aitken and Nucleation modes contributed more to the total particle concentration, while in the background site, it was Aitken and Accumulation modes. The annual average concentration in the urban and background sites was 2.5×10^4, 2.9×10^3 cm^{-3} respectively. Aitken-mode particles dominated both in the urban and background sites, having the lowest concentrations in summer (6.4×10^2 cm^{-3}) and monsoon period (1.09×10^2 cm^{-3}), respectively. The diurnal concentration depicts the role of transportation emissions in the urban site and the natural process of particle condensation in the background site. The long-range transportation of sources shows that during winter and autumn seasons, the westerly and north westerly wind brings biomass burning emissions smoke to urban region and it was seen less in the background conditions.

1. Introduction

Atmospheric aerosols persist in solid or liquid form (Kompalli et al., 2018; Sebastian et al., 2021), determine region's air quality, and contribute to climate change and human health (Ambade and Kurwadkar, 2021; Hussain et al., 2023; Trechera et al., 2023). These are known for its increased toxicity due to their higher surface area and volume. The aerosol concentration is quantified in terms of particle number concentration (PNC) per unit volume of air (particle number or # per cm^3) (Damayanti et al., 2023; Damayanti and Harrison, 2022). Recent research studies worldwide show that particle number concentration in urban and background regions is dominated by particles with size ranges <100 nm, i.e., often referred to as ultrafine particles, which attain peak concentration when the solar radiation increases (Van De Beek et al., 2021).

The most recent studies in Asian regions show that air quality in many Asian cities is deteriorating alarmingly and poses a major

Abbreviation: BG, Background; DMPS, Differential Mobility Particle Sizer; LDMA, Large Differential Mobility Analyser; LPM, Litre Per Minute; N_{acc}, Accumulation mode; N_{atk}, Aitken mode; NDL, New Delhi; nm, nanometre; N_{nuc}, Nucleation Mode; N_{tot}, Total Particles; ppm, parts per million; PNC, Particle Number Concentration; PNSD, Particle Number Size Distribution; $PM_{2.5}$, Particulate Matter with a diameter 2.5 or < 2.5 μm; RNC, Ranichauri; RH, Relative Humidity; SMPS+C, Scanning Mobility Particle Sizer with a Condensation Particle Counter; WS, Wind Speed.

* Corresponding author.
E-mail address: rajeevkumarmishra@dtu.ac.in (R.K. Mishra).

https://doi.org/10.1016/j.uclim.2024.102066
Received 14 May 2024; Received in revised form 27 June 2024; Accepted 16 July 2024
Available online 23 July 2024
2212-0955/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

threat to human health (Ambade, 2021; Ambade et al., 2020).

Delhi, being one of the megacities in the Asian continent, is not an exemption (Mohan et al., 2024; Rajagopal et al., 2023, 2024; Yadav et al., 2021, 2022b). Various measures have been taken, such as closing 1300 industries due to hazardous emissions, banning vehicles older than 15 years, and introducing compressed natural gas fuel-based vehicles to curb air pollution in Delhi. All the mitigation measures help to reduce the sulfur concentration to 50 ppm (Mishra et al., 2019; Pandey et al., 2022; Yadav et al., 2022a). A city like Delhi, where the $PM_{2.5}$ concentration increases steadily, consequently involves a significant increase in the particle number concentration from different sectors, such as automobile and industrial sectors (Ambade et al., 2022; Gani et al., 2021; Zheng et al., 2021). Around 80% of the population who reside in urban areas are exposed to levels of air pollution that exceed WHO recommendations (WHO, 2022), as is the case of the particles larger than 10 nm (1000 cm^{-3}/24-h mean and 20,000 cm^{-3}/1-h) (Su et al., 2021). Ultrafine particles are considered as a universal carrier of a wide variety of toxic chemicals to humans. The deposition of particles in the human respiratory system varies based on the size of the particles, such as nucleation, Aitken, and accumulation. Also, the concentration of these particles signifies the emissions sources in a particular region. These can enter the deeper regions of the respiratory system, such as alveoli, and can be translocated to the circulatory system and enter any organ of the body (Ambade, 2016; Austin et al., 2021; Gani et al., 2021; Pirhadi et al., 2020; Zhao et al., 2020).

This study analyzes the particle number size distribution (PNSD) and the variations of total and modal (nucleation, Aitken and accumulation) concentrations in two geographical locations, an urban and a background site, during different seasons. The concentration of pollutants in the urban and background sites is distinct due to the different types of sources present. Various studies presented the characteristics of these particles in different environmental conditions, but this study compares these two regions in the Indian domain. The main objectives of this work are to (i) investigate the variability of aerosol number size distributions in an urban and a background site and their modal particle number concentration, (ii) examine the diurnal variation of the modal particle number concentration and (iii) investigate the role of meteorological parameters in determining the concentration of these aerosols.

2. Methodology

2.1. Sampling sites

Two different regions were selected as sampling sites for this study. One is located in a highly urbanized area with a complex mixture of different sources, and the other is a background site where the anthropogenic sources are found in lesser concentration (Fig. 1 & Table. 1).

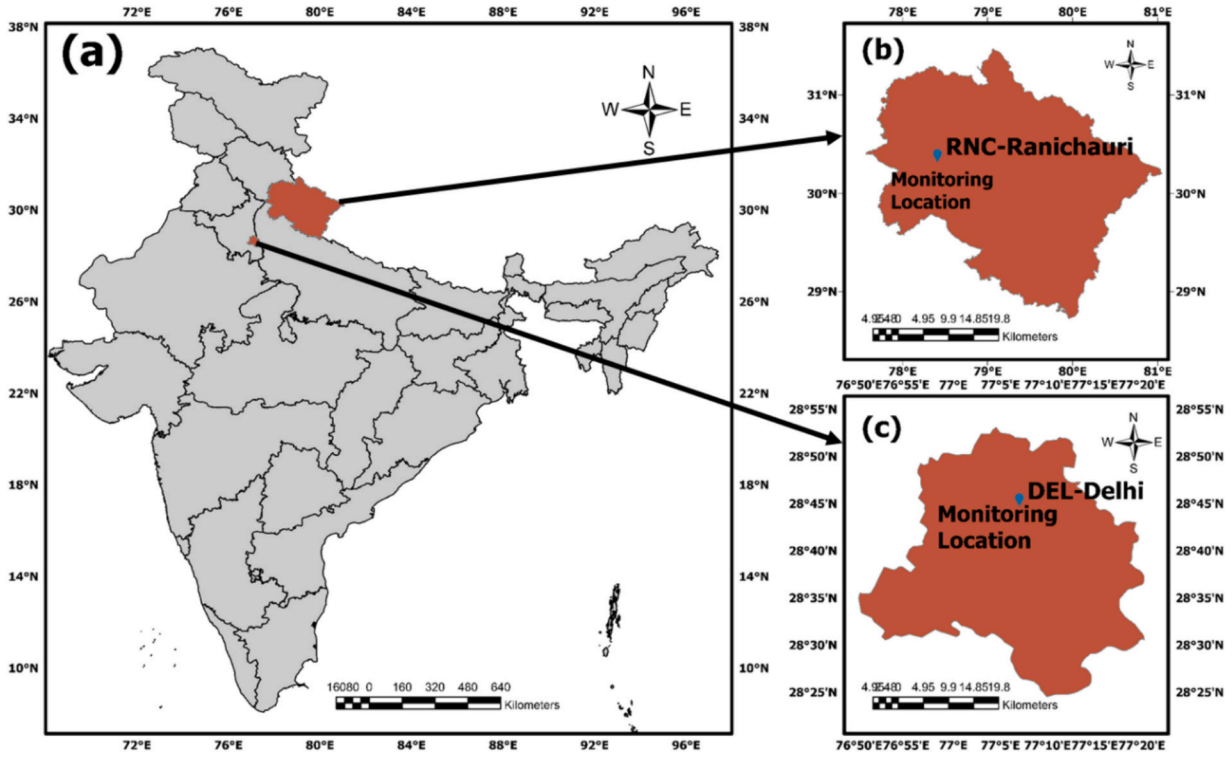


Fig. 1. (a) India Map, (b) Ranichauri (RNC) with a blue mark indicating the location of background site, and (c) Delhi (DEL) with a blue mark indicating the location of urban site. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.1.1. Urban site (Delhi)

The urban region observatory was in the northwest part of Delhi (28.75°N, 77.12°E). It is located inside the campus of Delhi Technological University, which is ~223 m above the mean sea level (amsl) and 8 m from the roadside. The region is highly populated, with commercial, residential, and industrial zones surrounding the sampling site. The sampling location is on the Bawana road connecting Delhi and Haryana, with a dense vehicular population of ~1300 vehicles per hour. During the sampling period, the meteorological conditions in the study area were temperature (25.89 ± 8.35 °C), wind speed (1.35 ± 0.83 m/s), and relative humidity (59.52 ± 18.73%), respectively.

2.1.2. Background site (Ranichauri)

The background observatory is in Ranichauri, Tehri Garhwal district in Uttarakhand state (30.25°N, 78.08°E). It is on the western Himalaya's southern slope, ~1930 m amsl. The nearest city from the background location is Rishikesh, which is 70 km away, and all other cities on different sides are above 100 km. Different types of lands, such as forest and agricultural activities, surround the study area. The meteorological condition in the study area during the sampling period is temperature (18.49 ± 7.02 °C), wind speed (2.39 ± 1.42 m/s) and relative humidity (58.67 ± 22.97%), respectively.

2.2. Instrumentation and campaign

Aerosol number size distributions were measured based on particle electrical mobility in an urban and a background site to compare them in terms of sources and meteorology (temperature, wind speed, wind direction, relative humidity). A GRIMM Inc. scanning mobility particle sizer model no. 5.403 measured particles from 10 to 1090 nm with a time interval of about ~7 min between each scan at the urban site. It comprises a long differential mobility analyser (LDMA) and a butanol vapor condensation particle counter. A β-radioactive source (Ni-63) was employed to bring the particles to an equilibrium charge distribution before they entered the classifier system. In the background location, a custom-built differential mobility particle sizer (DMPS) with a size range of 10 to 800 nm (30-size bins) is used for measurement. The Instrument is TSI make model no. 3772. It consists of a Vienna-type differential mobility analyser (DMA) that classifies charged particles based on their electrical mobility. The time interval between each scan in this instrument is ~10 min with a flow rate of 1 LPM and a sheath air flow rate of 5 LPM. The campaign was conducted in 2017 at the background site and 2022 at the urban site. The particles were then analysed using a scanning mobility particle sizer (SMPS) and a condensation particle counter (CPC). The data were analysed using R statistical software version 2023.12.1.

The meteorological parameters such as relative humidity, temperature, wind speed, and wind direction are also collected along with the particle number concentration in both locations to determine the role of meteorological parameters in different regions. The meteorological parameters are measured at a frequency of about 5 min. The accuracy of all the meteorological instruments follows the standards of the World Meteorological Organization. Accuracies of meteorological sensors are ±0.05 °C for temperature, ± 0.1 hPa for atmospheric pressure, ± 0.5 m/s, and ± 1° for wind speed and direction.

2.3. HYSPLIT trajectory

The long-range transportation of the pollutants analysis is done using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) open-source model developed by the Air Resource Laboratory (ARL), NOAA (National Oceanic and Atmospheric Administration). The seasonal average trajectories are plotted for 3 days (−72 h) back trajectory for every day (24 h, starting at 00 h) and the clustering is done for each season. The season has been considered for winter (December to March), summer (April to June), monsoon (July to September), and post-monsoon (October to November) for the years 2022 and 2017 of urban and background sites. The input of the models is taken from REANALYSIS data at a height of 1000 agl (above ground level) for both urban and background sites for analysing back trajectory in MeteoInfoMap.

2.4. Limitations of the study

The study measured particle number concentration of particles ranging from 10 to 800 nm in urban and background regions to compare the concentration of sources and type of particles present. Due to the limitations of the resources, the measurements are not done parallelly in both locations. Due to constraints in the instruments, two different instruments that work on the same principle are used for the measurements. These two are considered limitations of the present study, and still, the data used for analysis are checked for quality checks, and necessary statistical analysis was done to ensure the quality of the data. The significance of the data is tested using the Levene test and F test, and found the p-value <0.05, so the null hypothesis is rejected.

Table 1
Geographical locations of observation campaigns.

S·No	Station	Code	Latitude	Longitude	Geographical
1.	New Delhi	NDL	28.75°N	77.12°E	Urban (U)
2.	Ranichauri	RNC	30.25°N	78.08°E	Background (BG)

3. Result and discussion

3.1. Variation of modal particle number concentration in an urban and a background site

The modal particle number concentration, Nucleation (N_{nuc}, 10 to 30 nm), Aitken (N_{atk}, 30 to 100 nm), Accumulation (N_{acc}, 100 to 800 nm), and total particle number concentration (N_{tot}, < 800 nm) were compared between the two different sites (urban and background). The analysis shows N_{tot} concentration in the urban and background sites were found to be 2.5×10^4 cm^{-3} and 2.9×10^3 cm^{-3}, respectively (Table. 2). Trechera et al., 2023 reported 6 cities in Europe country showed higher PNC ranging from 1.35×10^4 to 1.05×10^4 # cm^{-3} near the traffic-related site whereas, cities from Italy showed lower concentrations in regional background ranges from 7.4×10^3 to 4.5×10^3 # cm^{-3}. Similar studies showed that the concentration in the urban site is higher compared with the background site (Rose et al., 2021; Trechera et al., 2023). The number concentration analysis reveals that in urban regions, the N_{atk} concentration was found to be higher than N_{nuc}, and followed by $N_{acc,}$ but in the background site, N_{atk} was higher than N_{acc}, followed by N_{nuc}, respectively. In urban regions, the higher concentration of N_{atk} and N_{nuc} is due to transport-related anthropogenic emissions, with an average of 5.7×10^2 cm^{-3}, 8.3×10^2 cm^{-3}, and 4.3×10^2 cm^{-3} for N_{nuc}, N_{atk}, and N_{acc}, similarly for the background site it was 0.4×10^2 cm^{-3}, 1.6×10^2 cm^{-3} and 0.9×10^2 cm^{-3}. On average, the difference in N_{tot} between the urban and background sites showed that N_{tot} was ~8 times higher in the urban site compared to the background one (Chatain et al., 2021).

The modal particle number concentration variations showed different patterns based on the characteristics and intensity of the sources in each study area. In both sites, N_{atk} showed a higher concentration than other modes, possibly due to the short-lived nature of N_{nuc} associated with the coagulation and condensation processes in the atmosphere (Chen et al., 2022; Sun et al., 2019). In the winter season, the maximum N_{tot} was observed, 1.0×10^3 cm^{-3} in the urban site and 2.3×10^2 cm^{-3} in background one (Şahin et al., 2022; Sun et al., 2019) (Fig. 2). The lowest concentration was observed for N_{atk} during the summer (6.4×10^2 cm^{-3}) in the urban site, while in background one was observed in the monsoon period 1.09×10^2 cm^{-3} (Table. 3). In summer is usual to find lesser concentration compared to winter due to the meteorology (Kalkavouras et al., 2020). The seasonal average for N_{tot} varies from one site to another. In the urban site, N_{tot} was higher in post-monsoon (3.0×10^4 cm^{-3}), followed by winter (2.9×10^4 cm^{-3}), monsoon (2.2×10^4 cm^{-3}), and summer (2.0×10^4 cm^{-3}) (Table. 3).

In the urban site, N_{nuc} and N_{atk} were similar during the summer and monsoon, but in the background site, the modal particle number concentration was as follows: $N_{atk} > N_{acc} > N_{nuc}$ for all seasons. The variations for N_{tot} between the urban and background sites were highest during the monsoon, which was ~13 times higher in urban regions, followed by post-monsoon ~12 times, winter ~7, and summer ~6.

3.2. Particle size distribution in an urban and a background site

Aerosol particle number size distribution was used as a tool to identify the sources of the particles. The size distribution pattern varies from the background to the urban site. The size distribution of the particles in urban site showed a single clear peak concentration between 80 and 90 nm, which typically refers to engine exhaust emissions (Chen et al., 2022; Kontkanen et al., 2020) (Fig. 3). In the background site, particle size distribution ranged gradually increase from 20 to 100 nm due to particle growth, a common phenomenon in this type of sites.

In the urban site, the PNC contribution from the transportation sector is more dominant than other sources (Chen et al., 2022; Gani et al., 2021). The highest concentration was obtained during winter, around 4×10^4 cm^{-3}, while the concentration ranged between 1-2×10^4 cm^{-3} in the rest of the seasons. Global studies showed similar behaviour during winter (Gani et al., 2021; Şahin et al., 2022). In the background site, the highest concentration was in winter, followed by summer, post-monsoon and monsoon.

In almost all the seasons, the gradual growth of the particles was observed. The size distribution pattern of the urban site showed that during all the seasons, the concentration of the particles emitted also lies in the range of 1 to 2×10^4 cm^{-3}, which is the result of direct emission from anthropogenic sources. In the background site, the sub-10 nm particles showed the particle concentration gradually reducing around 10 nm and increasing after 10 nm due to the change of particle phases from gaseous to solid due to the particle formation process. Consequently, the concentration of the size distribution varies from season to season. However, the emission profile remains similar in all the seasons, whereas in the urban site, the size distribution and emission profile change from season to season due to the influence of sources meteorology and the intensity of the sources.

Table 2
Average mean particle concentration in the urban and background sites during the study period.

	DEL-Urban Site				RNC-Background Site			
	N_{nuc}	N_{atk}	N_{acc}	N_{tot}	N_{nuc}	N_{atk}	N_{acc}	N_{tot}
Min	0	0	0	0	0	0	0	0
Max	10,338	5999	3422	159,194	2047	3573	2110	205,780
Mean	577	833	433	25,708	42	169	90	2968
SD	608	620	382	16,162	85	151	65	3277

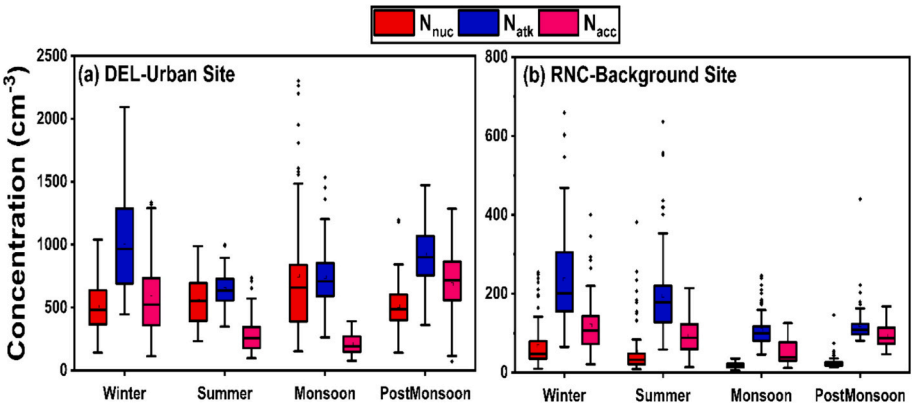


Fig. 2. Box whisker analysis of N_{nuc}, N_{atk}, and N_{acc} mode particles in an urban and a background site.

Table 3
Season-wise average statistics for modal particle number concentration in the urban and background sites during the study period.

		DEL-Urban Site				RNC-Background Site			
Season		N_{nuc}	N_{atk}	N_{acc}	N_{tot}	N_{nuc}	N_{atk}	N_{acc}	N_{tot}
Winter (Dec to Mar)	Min	10	75	42	2388	0	0	0	0
	Max	4097	5999	3422	140,484	1310	3573	2110	205,780
	Mean	510	1007	576	29,521	64	233	121	4145
	SD	497	834	448	19,359	106	197	88	5193
Summer (Apr to Jun)	Min	15	103	37	2906	3	36	9	511
	Max	4722	2253	1608	89,387	2047	1969	585	27,256
	Mean	562	645	269	20,280	54	196	93	3303
	SD	467	332	193	8995	110	153	49	2242
	Min	33	74	16	2461	2	19	3	247
	Max	10,338	4419	780	159,194	258	494	182	6993
	Mean	746	728	213	22,778	18	109	51	1728
Monsoon (Jul to Sep)	SD	878	423	117	15,762	19	60	32	899
	Min	0	0	0	0	6	38	28	863
	Max	3576	4298	3042	97,901	583	1632	974	23,682
	Mean	513	924	685	30,336	26	120	94	2479
Post Monsoon (Oct to Nov)	SD	520	592	411	15,331	45	102	43	1470

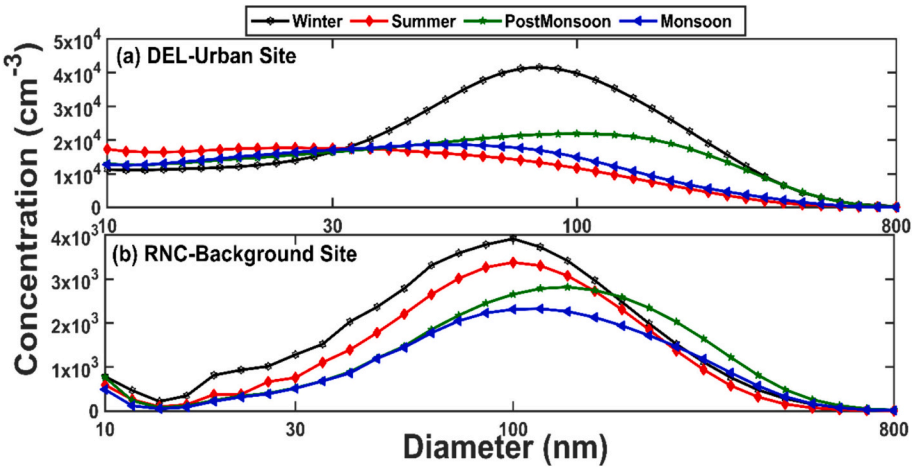


Fig. 3. Particle number size distribution of particles in (a) an urban and (b) a background site during different seasons.

3.3. Diurnal variations of modal particle number concentration in the urban and background sites

The diurnal variation of modal particle number concentration in the urban and background sites indicates how the concentration of the particles varies in terms of source during the different seasons.

In the urban site, N_{nuc} peaked from 8 to 12 h in the morning due to the fresh engine exhaust emissions, while in the background site during the winter and spring seasons, this peak was shifted (12 to 20 h) (Chen et al., 2022; Zhu et al., 2002, 2004). This was due to the increase in the formation of new particles into the atmosphere by gas-to-particle conversion induced by photochemistry and their subsequent growth. N_{atk} in the urban site showed a clear double hump model during the morning (8 to 11 h) and evening peak hours (16 to 20 h) which was associated with the transport of emissions from urban areas. Globally, this pattern signifies the role of pollution transportation in determining the particle concentration in the urban site (Fig. 4). The N_{acc} mode in the urban site followed a similar pattern to that of N_{atk}.

In the urban site, the N_{nuc} varied from 2.5×10^2 cm^{-3} to 1.5×10^3 cm^{-3}, while in the background site, it lies $<1 \times 10^3$ cm^{-3}, the evening peak higher due to the N_{acc} accumulation throughout the day. The N_{nuc} in monsoons was seen more during the morning hours due to the wet removal of the pre-existing larger particles, which promoted an increase in particles fresh from exhaust emission (1.5×10^3 cm^{-3}). The N_{atk} and N_{acc} mode particles showed higher concentration during the winter and post-monsoon periods. In the background site, the concentration of the three particle modes was higher during winter and summer, followed by post-monsoon and monsoon seasons.

3.4. Diurnal variation of meteorological conditions

The existing meteorological conditions influence particle number size distributions, mainly relative humidity and temperature. Thus, the relative humidity played a crucial role in inducing the particle to undergo condensation and coagulation processes (Mohan et al., 2024).

In the urban site, the lowest mean temperature was registered during winter (18 °C) and the highest during summer (33 °C), reaching a maximum of ~46 °C, while in the background site, it was only 38 °C (Fig. 5 & Table. 4). The relative humidity in the study sites was directly proportional to the particle concentration in the accumulation mode. The humidity >50% increased the possibility of

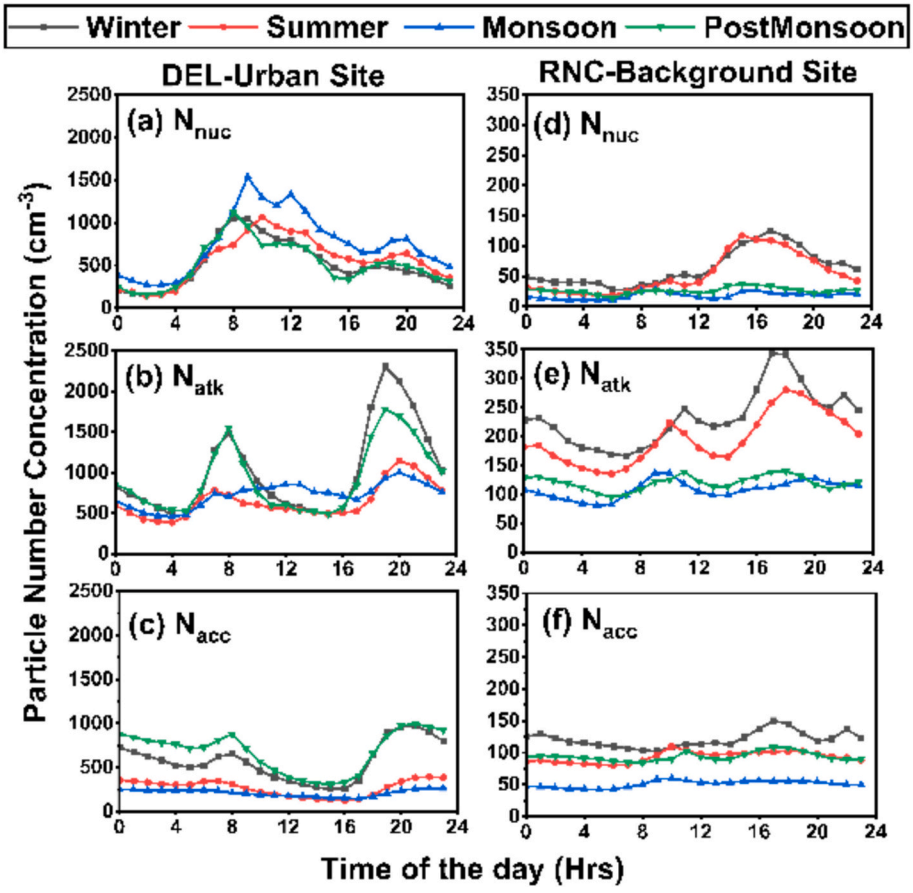


Fig. 4. Diurnal variation of modal particles number concentration in an urban and a background site during the different seasons.

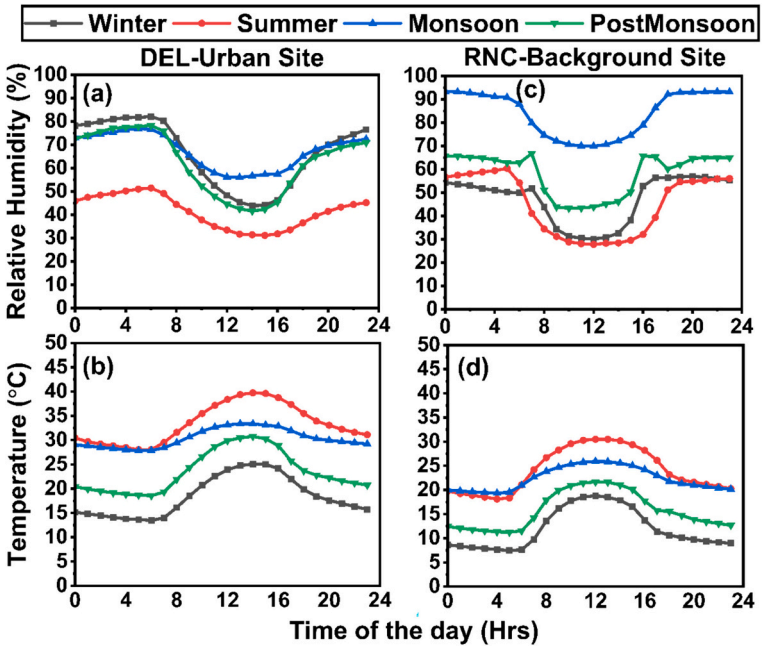


Fig. 5. Diurnal variation of relative humidity and temperature in the urban and background sites during the different seasons.

Table 4
Seasonal-wise average statistical summary of temperature (T), relative humidity (RH), and wind speed (WS) in the urban and background sites during the study period.

		DEL-Urban Site			RNC-Background Site		
Season		RH (%)	T (°C)	WS (m/s)	RH (%)	T (°C)	WS (m/s)
Winter	Min	22.6	6.9	0.1	13.6	0.1	0.1
	Max	95.1	40.3	5.6	100.0	32.6	7.8
	Mean	66.3	18.4	1.3	47.2	12.0	2.3
	SD	18.7	7.2	0.9	17.1	5.7	1.3
Summer	Min	20.8	19.5	0.2	8.6	9.0	0.1
	Max	88.3	46.3	9.1	98.6	38.2	8.9
	Mean	41.5	33.3	1.6	44.5	24.2	3.0
	SD	13.4	5.1	0.9	21.1	5.4	1.8
Monsoon	Min	38.0	23.6	0.2	52.9	16.2	0.1
	Max	88.5	38.5	4.3	100.0	29.3	7.6
	Mean	67.5	30.3	1.5	84.0	22.4	2.2
	SD	11.4	3.1	0.7	11.9	2.7	1.4
Post Monsoon	Min	29.4	11.5	0.1	25.1	5.7	0.1
	Max	90.8	35.8	3.4	88.4	26.0	5.0
	Mean	62.8	23.5	0.9	58.3	15.7	2.0
	SD	15.7	5.4	0.6	12.8	4.9	0.8

coagulation and condensation of the lower-size particles (N_{nuc} and N_{atk}) into accumulation mode particles, especially in the urban background. However, the geographical characteristics in the background promoted the arrival of air masses with a higher humidity than in the urban site, particle coagulation and condensation processes were inhibited by their low concentration. In the urban site, even >70% humidity increased the particle condensation and coagulation; consequently, the accumulation-mode particles also increased (Mohan et al., 2024; Zhu et al., 2006).

3.5. Role of wind speed and direction on particle dispersion

The wind speed and direction determine the pollutant's dispersion and rate from the sources to receptors. The urban site is surrounded by various buildings, which creates a shielding effect. Thus, the average wind speed was <2 m/s in all seasons, except for the post-monsoon season, with a similar wind direction in almost all seasons. In the post-monsoon season, the urban region receives an average wind speed throughout the season, ranging from 6 to 8 m/s. Similarly, the wind speed reached up to 10 m/s in the background site, but changes were found seasonally in the wind direction. Thus, particle dispersion occurs at a larger rate in this site during higher

wind speeds, and the particle concentration lies between 1.2 and 1.4 $\times$ 10^3 cm^{-3}.

The wind speed comparison between the urban and background sites showed a higher particle concentration in the background site, identifying an average wind speed of 2 to 3 m/s, while in the urban site, it was 1 to 2 m/s. The season-average wind speed was not >2 m/s in any season.

The mean N_{tot} may also increase based on the dispersion of pollutants from one region to another region due to the wind speed and direction. Calm winds (< 2 m/s) led to less particle dispersion, oppositely with winds >4 m/s (Fig. 6), being pollutants transported to the leeward side from the windward side. Other authors have also stated similar kind of findings in their study (Mohan and Mishra, 2022; Şahin et al., 2022).

3.6. Long-range dispersion pattern in urban and background site

The trajectory plot shows that the pollutants are not only generated from the local sources but also transported from long-range transportation. To analyse the impact of long-range transportation of pollutants from different faraway regions, the back trajectories were analysed for 2022 and 2017 in urban and background regions (Fig. 7&8). In the urban region of Delhi, the back trajectory shows that for most of the seasons, the winds are blown from the western side of the study location (Sharma et al., 2022). During the onset of winter, and post-monsoon season (October to March), most of the wind comes from the neighbouring states, namely Punjab and Haryana (~60 to 80%). This period is known for the crop residue burning period, where burning the remaining crop wastes brings large amounts of smoke particles to Delhi and causes smog formation. Severe air pollution events are attributed to the downwind movement of aerosols from biomass burning (Kanawade et al., 2020). In our study, the concentration of ultrafine particles was higher during the mentioned period. Around 82% of the winds come from the same direction possible, causing long-range transportation of aerosols apart from the local sources. Similarly, during summer, the winds are from Rajasthan, i.e., the westerly and north-westerly direction brings a significant amount of dust from the Thar desert that contributed to concentrations during the summer period (Banoo et al., 2022; Singh et al., 2020). However, from July to September, the wind pattern in the urban regions changes and the winds travel from the Arabian Sea or Bay of Bengal, i.e., from the southwestern coast. These winds are responsible for the precipitation in both the study regions, bringing many marine aerosols to the study location. The shift in the dispersion of wind patterns could attributed to the dominant monsoon winds from the Indian Ocean. In the background region, the wind pattern is more or less similar to the urban regions. The fascinating thing about the background region is that the hilly terrain restricts long-range transportation. At the same time, the regional wind sources contribute to around 50% of the wind in almost all the seasons in the background conditions (Fig. 8). This analysis shows that the concentration of ultrafine particles is based on the long-range transportation of the winds and associated pollutants. Also, it varies from one geographical domain to another geographical condition, such as urban and background conditions.

4. Conclusions

This study analysed the variation in aerosol particle number concentration in urban and background conditions. The annual

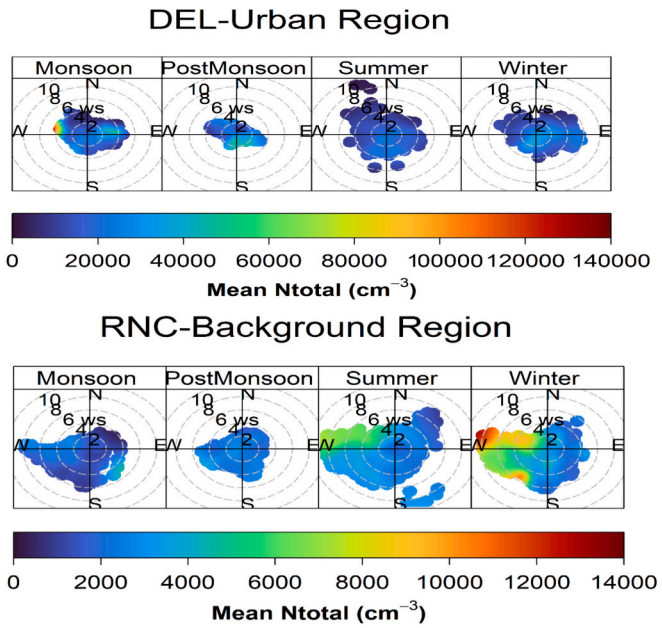


Fig. 6. Polar cluster analysis of seasonal total particle concentration in the urban and background sites during different seasons.

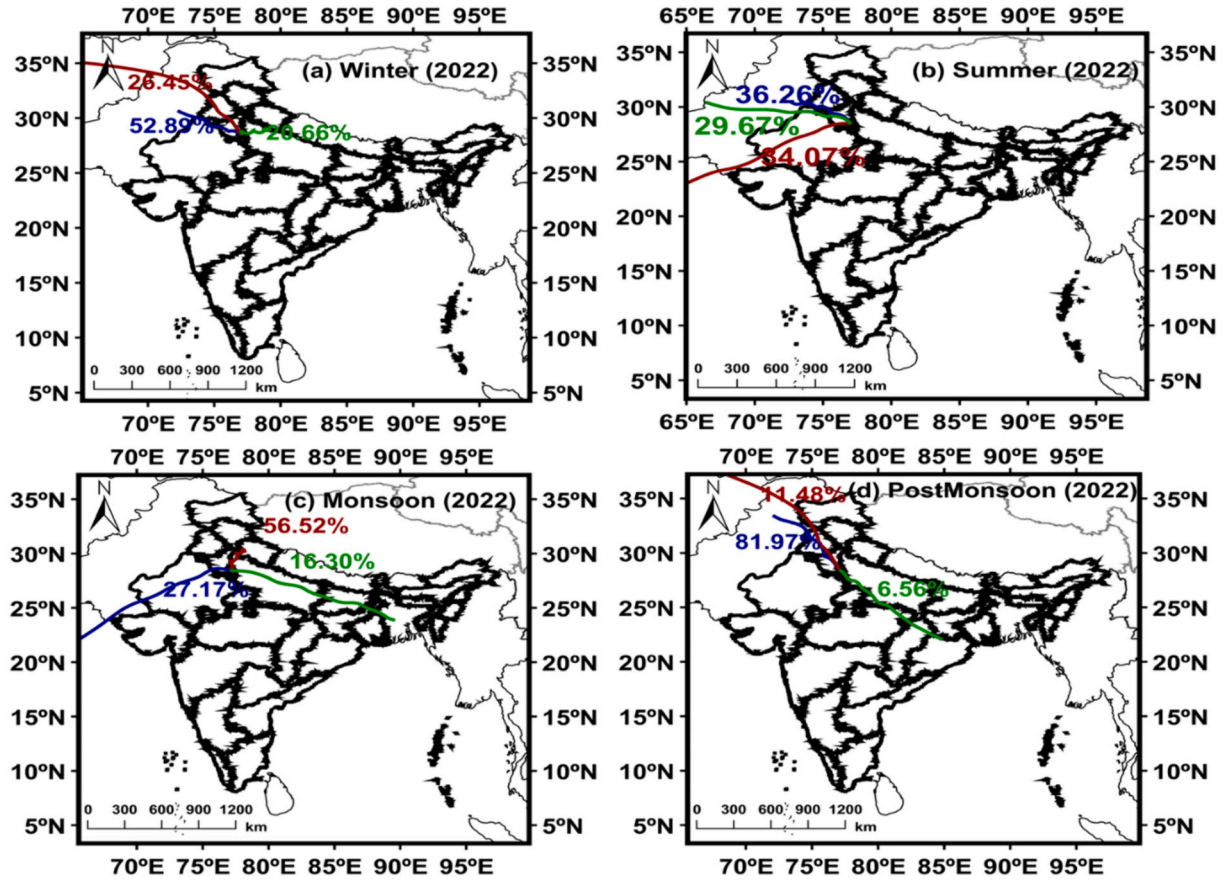


Fig. 7. Back trajectory plots in urban site for the different seasons.

average N_{tot} in the urban site was 2.5×10^4 cm^{-3}, and the concentration in the background site was 2.9×10^3 cm^{-3}. Particle number concentration in the urban site was higher for N_{atk} followed by N_{nuc} and N_{acc}, whereas, in the background site, it was for N_{atk} followed by N_{nuc} and N_{acc}. In both urban and background locations, the concentration of N_{atk} was found more in urban and background regions. The lowest concentration (6.4×10^2 cm^{-3}) for N_{atk} in the urban site was observed during the summer, and in the background site during the monsoon season (1.09×10^2 cm^{-3}). The urban regions with more anthropogenic emissions, especially the transportation sources, exhibit a peak around 80 to 90 nm. Similarly, in the background site, the size distribution changes gradually from lower to higher sizes due to the particle growth.

The diurnal variation of the particles in the urban site and background regions shows the impact of anthropogenic sources, where transport emissions are higher during morning and evening peak hours. In the background site, particle concentration was high during the evening hour due to their accumulation from the morning emissions to the evening in the study area. The meteorological conditions promoted the secondary formation of aerosol, which was observed when the RH was >70% in the urban site and > 50% in the background site. The annual wind speed in the urban site was <2 m/s and ranged from 6 to 10 m/s in the background site. The back-trajectory analysis shows that in both the study sites, during different seasons, wind pattern changes, especially the global wind direction, which contributed to different types of particles originating from various sources such as burning, sand dust and marine aerosols. The comparative study in urban and background sites provides a detailed assessment of the role of anthropogenic emission in determining the particle number concentration in the study regions.

The study helps to identify the role of different sources, meteorology and geographical condition in determining the concentration, size distribution, influence of other factors which are crucial to understand the dynamics of the ultrafine particles in different environmental conditions. Further similar studies can be conducted in different environmental condition to examine the sources, concentration and potential health impacts caused due to these UFP's. Urban regions pose a serious threat to exposure to these particles. The deposition potential of the particles in the respiratory system is based on the concentration and properties of the particles, which increases toxicity. The present study may be useful to estimate the long-term and short-term seasonal exposure. The measured concentration can be further used for simulating health impacts such as pulmonary deposition. The study also emphasizes the need to initiate mitigation measures to improve air quality and policy-driven action measures to attain sustainability.

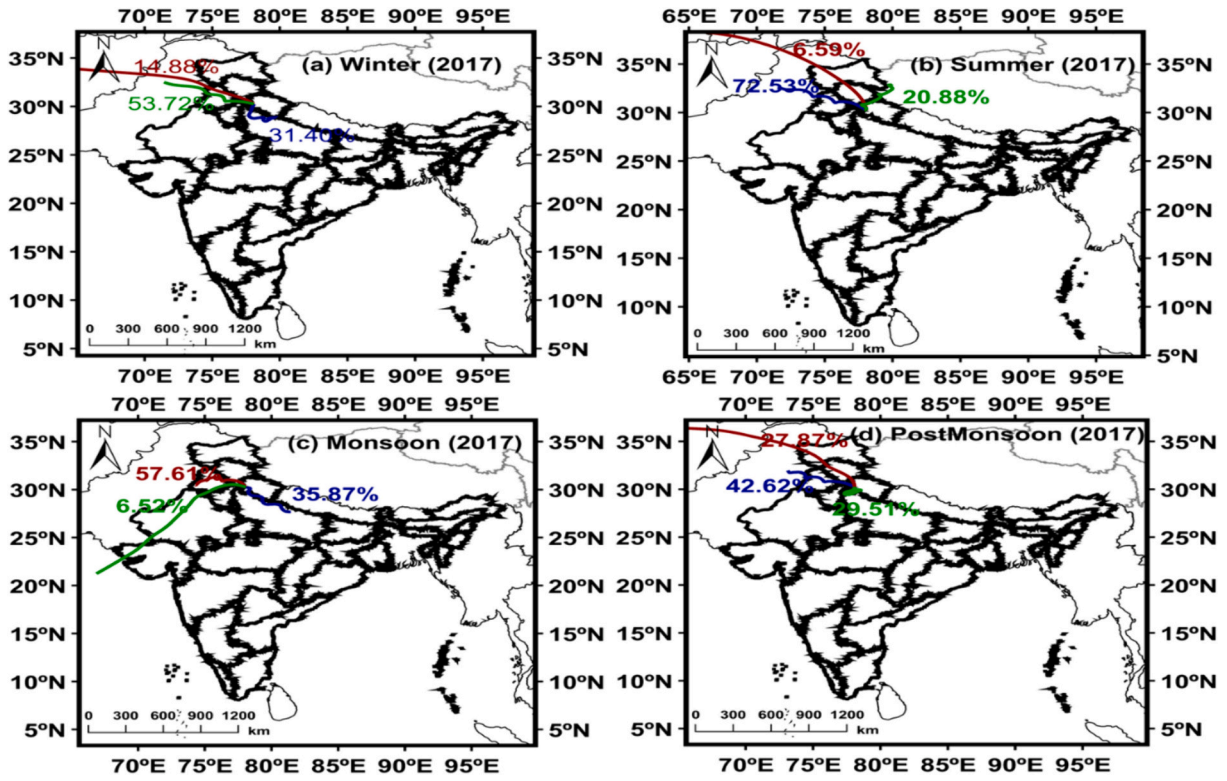


Fig. 8. Back trajectory plots in background site for the different seasons.

Submission, declaration, and verification

The work has not been published anywhere previously in any form, and it is not in consideration for publication elsewhere. All authors approve of the publication.

Credit statement

Vignesh Mohan – Data collection, Conceptualization, Visualization, formal analysis, reviewing, and editing.
Vijay Kumar Soni – Discussion, reviewing, and editing.
Rajeev Kumar Mishra – Investigation, discussion, reviewing, and editing.

Funding

Not applicable.

Code availability

Not applicable.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent to publish

All the authors approved the manuscript for publication.

Ethical responsibilities of authors

All authors have read, understood, and have complied as applicable with the statement on "Ethical Responsibilities of Authors" as found in the Instructions for Authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on reasonable request.

Acknowledgement

The authors thank Advanced Air and Acoustics Research Laboratory members, the Department of Environmental Engineering, Delhi Technological University (DTU), and India Meteorological Department (IMD) for providing instrumental facilities to conduct this study.

References

Ambade, B., 2016. Characterization of PM10 over urban and rural sites of Rajnandgaon, Central India. Nat. Hazards 80 (1), 589–604. https://doi.org/10.1007/s11069-015-1985-2.

Ambade, B., 2021. Emission sources, Characteristics and risk assessment of particulate bound Polycyclic Aromatic Hydrocarbons (PAHs) from traffic sites, pp. 1–15.

Ambade, B., Kurwadkar, S., 2021. Emission reduction of black carbon and polycyclic aromatic hydrocarbons during COVID-19 pandemic lockdown, pp. 1081–1095.

Ambade, B., Sankar, T.K., Kumar, A., Sethi, S.S., 2020. Characterization of PAHs and n-alkanes in atmospheric aerosol of Jamshedpur City, India, 24 (2), 1–11. https://doi.org/10.1061/(ASCE)HZ.2153-5515.0000490.

Ambade, B., Sankar, T.K., Sahu, L.K., Dumka, U.C., 2022. Understanding sources and composition of black carbon and PM2.5 in urban environments in East India. Urban Sci. 6 (3) https://doi.org/10.3390/urbansci6030060.

Austin, E., Xiang, J., Gould, T.R., Shirai, J.H., Yun, S., Yost, M.G., Larson, T.V., Seto, E., 2021. Distinct ultrafine particle profiles associated with aircraft and roadway traffic. Environ. Sci. Technol. 55 (5), 2847–2858. https://doi.org/10.1021/acs.est.0c05933.

Banoo, R., Sharma, S.K., Vijayan, N., Mandal, T.K., 2022. Assessment of potential source and the source region of particulate matter in an urban area of Delhi, India. Aerosol Sci. Eng. 6 (3), 231–245. https://doi.org/10.1007/s41810-022-00139-4.

Chatain, M., Alvarez, R., Ustache, A., Rivière, E., Favez, O., Pallares, C., 2021. Simultaneous roadside and urban background measurements of submicron aerosol number concentration and size distribution (in the range 20–800 nm), along with chemical composition in Strasbourg, France. Atmosphere 12 (1). https://doi.org/10.3390/ATMOS12010071.

Chen, Y., Masiol, M., Squizzato, S., Chalupa, D.C., Zíková, N., Pokorná, P., Rich, D.Q., Hopke, P.K., 2022. Long-term trends of ultrafine and fine particle number concentrations in New York state: apportioning between emissions and dispersion. Environ. Pollut. 310 (July) https://doi.org/10.1016/j.envpol.2022.119797.

Damayanti, S., Harrison, R.M., 2022. Long-term trends of Total Number Count and Black Carbon at Rural, Urban Background, and Urban Traffic Sites in the UK (September).

Damayanti, S., Harrison, R.M., Pope, F., Beddows, D.C.S., 2023. Limited impact of diesel particle filters on road traffic emissions of ultrafine particles. Environ. Int. 174 (March), 107888 https://doi.org/10.1016/j.envint.2023.107888.

Gani, S., Chambliss, S.E., Messier, K.P., Lunden, M.M., Apte, J.S., 2021. Spatiotemporal profiles of ultrafine particles differ from other traffic-related air pollutants: lessons from long-term measurements at fixed sites and mobile monitoring. Environ. Sci. Atmosph. 1 (7), 558–568. https://doi.org/10.1039/d1ea00058f.

Hussain, A.J., Sankar, T.K., Vithanage, M., Ambade, B., Gautam, S., 2023. Black carbon emissions from traffic contribute sustainability to air pollution in urban cities of India. Water Air Soil Pollut. 234 (4) https://doi.org/10.1007/s11270-023-06232-9.

Kalkavouras, P., Bougiatioti, A., Grivas, G., Stavroulas, I., Kalivitis, N., Liakakou, E., Gerasopoulos, E., Pilinis, C., Mihalopoulos, N., 2020. On the regional aspects of new particle formation in the eastern Mediterranean: a comparative study between a background and an urban site based on long term observations. Atmos. Res. 239 (August 2019), 104911 https://doi.org/10.1016/j.atmosres.2020.104911.

Kanawade, V.P., Srivastava, A.K., Ram, K., Asmi, E., Vakkari, V., Soni, V.K., Varaprasad, V., Sarangi, C., 2020. What caused severe air pollution episode of November 2016 in New Delhi? Atmos. Environ. 222 (November 2019), 117125 https://doi.org/10.1016/j.atmosenv.2019.117125.

Kompalli, S.K., Babu, S.S., Udayasoorian, C., Jayabalakrishnan, R.M., 2018. Role of anthropogenic emissions and meteorology on ultrafine particle bursts over a high altitude site in Western Ghats during pre-monsoon. J. Atmos. Sol. Terr. Phys. 179, 378–388. https://doi.org/10.1016/j.jastp.2018.09.001.

Kontkanen, J., Deng, C., Fu, Y., Dada, L., Zhou, Y., Cai, J., Daellenbach, K.R., Hakala, S., Kokkonen, T.V., Lin, Z., Liu, Y., Wang, Y., Yan, C., Petäjä, T., Jiang, J., Kulmala, M., Paasonen, P., 2020. Size-resolved particle number emissions in Beijing determined from measured particle size distributions. Atmos. Chem. Phys. 20 (19), 11329–11348. https://doi.org/10.5194/acp-20-11329-2020.

Mishra, R.K., Pandey, A., Pandey, G., Kumar, A., 2019. The effect of odd-even driving scheme on PM 2.5 and PM 1.0 emission. Transp. Res. Part D: Transp. Environ. 67, 541–552. https://doi.org/10.1016/j.trd.2019.01.005.

Mohan, V., Mishra, R.K., 2022. A picture of Delhi's regional air quality during diminished anthropogenic activities in the COVID-19 era. Arab. J. Geosci. 15 (15), 1–14. https://doi.org/10.1007/s12517-022-10567-8.

Mohan, V., Kumar Soni, V., Kumar Mishra, R., 2024. Analysing the impact of day-night road traffic variation on ultrafine particle number size distribution and concentration at an urban site in the megacity Delhi. Atmos. Pollut. Res. 15 (4), 102065 https://doi.org/10.1016/j.apr.2024.102065.

Pandey, A., Mishra, R.K., Pandey, G., 2022. Assessing the Effect of Vehicle Traits on CO and HC emmisions During Different Engine Iddling Speeds, 5, pp. 41–44 doi: IJMPE-IRAJ-DOIONLINE-18669.

Pirhadi, M., Mousavi, A., Sowlat, M.H., Janssen, N.A.H., Cassee, F.R., Sioutas, C., 2020. Relative contributions of a major international airport activities and other urban sources to the particle number concentrations (PNCs) at a nearby monitoring site. Environ. Pollut. 260, 114027 https://doi.org/10.1016/j.envpol.2020.114027.

Rajagopal, K., Ramachandran, S., Mishra, R.K., 2023. Roadside measurements of nanoparticles and their dynamics in relation to traffic sources in Delhi: impact of restrictions and pollution events. Urban Clim. 51 (July), 101625 https://doi.org/10.1016/j.uclim.2023.101625.

Rajagopal, K., Mohan, V., Kumar, R., 2024. Are Delhi residents exposed to lesser particle number concentration due to the firework ban in the city? Air Qual. Atmos. Health. https://doi.org/10.1007/s11869-024-01532-3, 0123456789.

Rose, C., Collaud Coen, M., Andrews, E., Lin, Y., Bossert, I., Lund Myhre, C., Tuch, T., Wiedensohler, A., Fiebig, M., Aalto, P., Alastuey, A., Alonso-Blanco, E., Andrade, M., Artinano, B., Arsov, T., Baltensperger, U., Bastian, S., Bath, O., Beukes, J.P., Laj, P., 2021. Seasonality of the particle number concentration and size distribution: a global analysis retrieved from the network of global atmosphere watch (GAW) near-surface observatories. Atmos. Chem. Phys. 21 (22), 17185–17223. https://doi.org/10.5194/acp-21-17185-2021.

Şahin, Ü.A., Harrison, R.M., Alam, M.S., Beddows, D.C.S., Bousiotis, D., Shi, Z., Crilley, L.R., Bloss, W., Brean, J., Khanna, I., Verma, R., 2022. Measurement report: interpretation of wide-range particulate matter size distributions in Delhi. Atmos. Chem. Phys. 22 (8), 5415–5433. https://doi.org/10.5194/acp-22-5415-2022.

Sebastian, M., Kanawade, V.P., Soni, V.K., Asmi, E., Westervelt, D.M., Vakkari, V., Hyvärinen, A.P., Pierce, J.R., Hooda, R.K., 2021. New particle formation and growth to climate-relevant aerosols at a background remote site in the Western Himalaya. J. Geophys. Res. Atmos. 126 (7), 1–16. https://doi.org/10.1029/2020JD033267.

Sharma, P., Peshin, S.K., Soni, V.K., Singh, S., Beig, G., Ghosh, C., 2022. Seasonal dynamics of particulate matter pollution and its dispersion in the city of Delhi, India. Meteorog. Atmos. Phys. 134 (2), 1–18. https://doi.org/10.1007/s00703-021-00852-8.

Singh, J., Gupta, P., Gupta, D., Verma, S., Prakash, D., Payra, S., 2020. Fine particulate pollution and ambient air quality: a case study over an urban site in Delhi, India. J. Earth Syst. Sci. 129 (1) https://doi.org/10.1007/s12040-020-01495-w.

Su, S., Lv, T., Lai, Y., Mu, J., Ge, Y., Giechaskiel, B., 2021. Particulate emissions of heavy duty diesel engines measured from the tailpipe and the dilution tunnel. J. Aerosol Sci. 156 (March), 105799 https://doi.org/10.1016/j.jaerosci.2021.105799.

Sun, J., Birmili, W., Hermann, M., Tuch, T., Weinhold, K., Spindler, G., Schladitz, A., Bastian, S., Löschau, G., Cyrys, J., Gu, J., Flentje, H., Briel, B., Asbach, C., Kaminski, H., Ries, L., Sohmer, R., Gerwig, H., Wirtz, K., Wiedensohler, A., 2019. Variability of black carbon mass concentrations, sub-micrometer particle number concentrations and size distributions: results of the German ultrafine aerosol network ranging from city street to high alpine locations. Atmos. Environ. 202 (December 2018), 256–268. https://doi.org/10.1016/j.atmosenv.2018.12.029.

Trechera, P., Garcia-Marlès, M., Liu, X., Reche, C., Pérez, N., Savadkoohi, M., Beddows, D., Salma, I., Vörösmarty, M., Casans, A., Casquero-Vera, J.A., Hueglin, C., Marchand, N., Chazeau, B., Gille, G., Kalkavouras, P., Mihalopoulos, N., Ondracek, J., Zikova, N., Querol, X., 2023. Phenomenology of ultrafine particle concentrations and size distribution across urban Europe. Environ. Int. 172 (January) https://doi.org/10.1016/j.envint.2023.107744.

Van De Beek, E., Kerckhoffs, J., Hoek, G., Sterk, G., Meliefste, K., Gehring, U., Vermeulen, R., 2021. Spatial and spatiotemporal variability of regional background ultrafine particle concentrations in the Netherlands. Environ. Sci. Technol. 55 (2), 1067–1075. https://doi.org/10.1021/acs.est.0c06806.

WHO, 2022. World Health Statistics. World Health, pp. 1–177.

Yadav, S.K., Kompalli, S.K., Gurjar, B.R., Mishra, R.K., 2021. Aerosol number concentrations and new particle formation events over a polluted megacity during the COVID-19 lockdown. Atmos. Environ. 259 (May), 118526 https://doi.org/10.1016/j.atmosenv.2021.118526.

Yadav, S.K., Mishra, R.K., Gurjar, B.R., 2022a. Assessment of the effect of the judicial prohibition on firecracker celebration at the Diwali festival on air quality in Delhi, India. Environ. Sci. Pollut. Res. 29 (57), 86247–86259. https://doi.org/10.1007/s11356-021-17695-w.

Yadav, S.K., Mishra, R.K., Gurjar, B.R., 2022b. Ultrafine particle number concentration and its size distribution during Diwali festival in megacity Delhi, India: Are 'green crackers' safe? J. Environ. Manag. 317 (July 2021), 115459 https://doi.org/10.1016/j.jenvman.2022.115459.

Zhao, J., Birmili, W., Wehner, B., Daniels, A., Weinhold, K., Wang, L., Merkel, M., Kecorius, S., Tuch, T., Franck, U., Hussein, T., Wiedensohler, A., 2020. Particle mass concentrations and number size distributions in 40 homes in Germany: indoor-to-outdoor relationships, diurnal and seasonal variation. Aerosol Air Qual. Res. 20 (3), 576–589. https://doi.org/10.4209/aaqr.2019.09.0444.

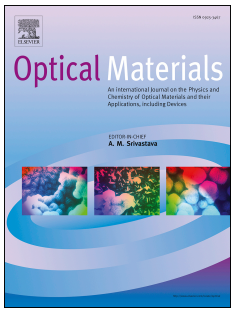
Zheng, T., Wang, H.W., Li, X.B., Peng, Z.R., He, H.Di., 2021. Impacts of traffic on roadside particle variations in varied temporal scales. Atmos. Environ. 253 (October 2020), 118354 https://doi.org/10.1016/j.atmosenv.2021.118354.

Zhu, Y., Hinds, W.C., Kim, S., Shen, S., Sioutas, C., 2002. Study of ultrafine particles near a major highway with heavy-duty diesel traffic. Atmos. Environ. 36 (27), 4323–4335. https://doi.org/10.1016/S1352-2310(02)00354-0.

Zhu, Y., Hinds, W.C., Shen, S., Sioutas, C., 2004. Seasonal trends of concentration and size distribution of ultrafine particles near major highways in Los Angeles. Aerosol Sci. Technol. 38 (Suppl. 1), 5–13. https://doi.org/10.1080/02786820390229156.

Zhu, Y., Kuhn, T., Mayo, P., Hinds, W.C., 2006. Comparison of daytime and nighttime concentration profiles and size distributions of ultrafine particles near a major highway. Environ. Sci. Technol. 40 (8), 2531–2536. https://doi.org/10.1021/es0516514.

Journal Pre-proof



Ge-Se-Te based Penrose Photonic Quasi-Crystal fiber for SCG covering 2-21 μm MIR Regime

Akash Khamaru, Ajeet Kumar

PII: S0925-3467(24)01032-2

DOI: https://doi.org/10.1016/j.optmat.2024.115849

Reference: OPTMAT 115849

To appear in: *Optical Materials*

Received Date: 27 June 2024

Revised Date: 11 July 2024

Accepted Date: 17 July 2024

Please cite this article as: A. Khamaru, A. Kumar, Ge-Se-Te based Penrose Photonic Quasi-Crystal fiber for SCG covering 2-21 μm MIR Regime, *Optical Materials*, https://doi.org/10.1016/j.optmat.2024.115849.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier B.V.

Journal Pre-proof

Ge-Se-Te based Penrose Photonic Quasi-Crystal fiber for SCG covering 2-21 μm MIR Regime

Akash Khamaru[a], Ajeet Kumar[a,*]

[a]**Advanced Photonics Simulation Research Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, Shahbad Daulatpur, Delhi 110042, India**

[*]**Corresponding Author:** ***ajeetdph@dtu.ac.in***

Abstract:

This work proposes a Penrose-based photonic quasi-crystal fiber (PQF) composed of Arsenic-free, non-toxic chalcogenide glass for Supercontinuum (SC) generation consisting of quasi-crystal lattice of air holes having eight-fold rotational symmetry. Following optimising the structural parameters, the proposed PQF offers zero dispersion wavelength (ZDW) at around 3.8 μm, an effective mode area of 30 μm^2, and a respective high nonlinearity of 824 $W^{-1}Km^{-1}$ at 2 μm wavelength. By injecting a secant hyperbolic optical pulse having a central wavelength of 4 μm with 20kW peak power and pulse width of 50 fs, a broadband SC of 2-21 μm is generated, which mostly covers the transparency range of the proposed material as well as Mid-IR regime. To the best of our understanding, this particular type of PQF for broadband SC generation has been proposed for the first time, serving as the novelty factor. Apart from novelty, the wide-band SC assists the proposed PQF in becoming a better counterpart for applications like optical coherence tomography (OCT), frequency metrology, and other Mid-IR applications.

1. Introduction:

Supercontinuum generation (SCG) is the phenomenon in which a short, high-intensity optical pulse transforms into a broadband spectrum having a range of wavelengths when it travels down through a nonlinear optical medium such as optical fibers. The transformation takes place due to the interaction of optical pulse with the nonlinear and dispersive effects such as stimulated Brillouin scattering (SBS) [1], Stimulated Raman Scattering (SRS) [1, 2], self-steepening, four-wave mixing (FWM) [2, 3], self-phase and cross-phase modulation (SPM-XPM) [4, 5], and many more. The variation in the effect of nonlinearities depends on the structural and material composition of the utilized medium for SC generation. Micro-structured holey fibers (MOFs) are superior candidates over conventional fibers due to certain advantages like endless single-mode operation, high power delivery, high birefringence and better dispersion adjustment [6-12]. MOFs like photonic crystal fiber (PCF) [13, 14], photonic bandgap fibers (PBGF) [15], photonic quasi-crystal fibers (PQF) [16, 17], and micro-porous core fibers (M-PCF) [18] are successful candidates for enhancing the nonlinear and dispersive outcomes and, more importantly, for the highly broadband SC generation. The phenomenon of supercontinuum was demonstrated by Alfano and Shapiro for the first time in 1970 [19]. In 2000, Ranka et al. were the first to generate SC spectra in PCF [20]. Afterward, works related to SCG in PCF continued to flourish [21-25]. Various notable works about SC generation in PCF have recently come to light. In 2024, Wen et al. generated a highly coherent SC spectrum spanning from 1300-2800 nm at low peak power of 1000 W [26]. In the same year, Nguyen

Thi et al. achieved a Supercontinuum spectrum spanning from 0.843 μm to 2.517 μm at the pumping of 1.57 μm with a peak power of 350 W [27].

Apart from PCF, PQFs have started to grab attention by emphasizing their features like ultra-flattened dispersion [28, 29], large modal area [30], endless single-mode operation due to the large magnitude of the cut-off ratio, and many more. The air holes do not exhibit any recurring translational pattern but attain two different types of lattice units, which repeat themselves by maintaining the rotational symmetry across the cross-section of the fiber. These characteristics allow PQFs to have greater structural degrees of freedom and distinctive cladding configurations that provide greater guiding feature tuning flexibility [31]. In the field of SCG, PQF also stands out as a competitive candidate that can deliver broadband SC spectrum at comparable peak powers and a short length of fiber. In 2016, Ahmad et al. proposed a circular lattice As_2Se_3-based PCF, which can be christened as sunflower-type PQF, has the ability to display a wide band of SC, i.e., 1.2-9.3 μm at only 0.5 cm of length [32]. Vallliammai et al. designed a circular PQF the following year and generated an SC spanning 2-15 μm with a very low confinement loss [33]. In the same year, Zhao et al. extracted an SC band of 1-12.5 μm by numerically simulating a twin bow-type PQF at the pumping of 2.94 μm operating wavelength [34]. Other than SCG, In 2014, Su et al. introduced twin-bow type PQF for high birefringence offering high magnitude (~2.43×10-3) [16]. In the same year, an octagonal penrose-based double cladding PQF was proposed by Liu et al., which delivers a very low loss profile, and ultra-flattened dispersion [17]. From the past literature discussion, it is clear that Penrose-based PQF has already been proposed for EMA and dispersion engineering calculations. Therefore, to upscale the research focussing on Penrose-based PQF, We have selected that specific PQF design for the Supercontinuum generation, which is proposed for the first time in the field of SCG applications and introduces a difference between the past works and our presented work.

Higher energy in the near-IR (and visible) regions excites overtone/harmonic and combination bands, whereas lower energy mid-IR radiation excites the fundamental vibration bands. These fundamental vibrations are more powerful than the overtone and combination bands that occur from the fundamental vibrations because they originate from the molecule's most basic vibrational states. This leads to significantly cleaner spectra originating from fundamental vibrations, which produce a distinct fingerprint for the material that may be utilized for identification. This respective reason increases the importance of mid-IR spectroscopy. Silica-based fibers can only offer SC band in the near-IR region since beyond 2.5 μm, high material absorption comes into play, which disrupts the respective SC band. Chalcogenide glasses and their various compositions inherit transparency beyond the near-IR region and wrap the mid-IR region [35], which qualifies them as the perfect choice for PCF fabrication. Along with a large span of transparency, chalcogenide glasses inherit a high magnitude of linear and nonlinear refractive indices. Tellurium (Te)-based chalcogenides can offer a transparency range beyond 25 μm. $Ge_{20}Se_{70}Te_{10}$ is a representative of Ge-Se-Te (GST) glass, which acquires high nonlinearity of order of 10^{-18} $W^{-1}km^{-1}$ along with a transmission range of 2-25 μm [40]. Apart from these qualities, GST glass has also been proven to be non-toxic due to the absence of Arsenic in molecular compositions [36]. The past works prove that the particular glass composition was readily used for designing and fabricating PCF structures for SCG applications due to its numerous advantages. In 2019, Lemiere et al. fabricated a step-index GST-based fiber and successfully generated an SC spectrum of 2-14 μm [37]. In 2021, a

broadband SC band of 1.7 to 18 μm was achieved by simulating a GST-based step-index fiber with the input of 200 kW peak power pulse centered at 8.15 μm wavelength [38]. Since $Ge_{20}Se_{70}Te_{10}$ has several advantages, we have chosen this glass to design our PQF structure.

In this paper, we propose a novel penrose-type PQF structure that consists of $Ge_{20}Se_{70}Te_{10}$ as background material and is surrounded by air holes. The air holes are located in such a manner that they do not exhibit translational symmetry but exhibit an eight-fold rotational symmetry. The proposed fiber is employed for generating SC spectra, which deliver a wide-band SC of 2-21 μm at the input of 4 μm pump wavelength with a peak power of 20 kW and respective fiber length of 4 mm. Apart from wide-band SC, the devised fiber has the capability to offer high nonlinearity of 824 $W^{-1}Km^{-1}$ at 2 μm wavelength. Due to high nonlinearity and wide transmission window, the PQF proves its potential to become a better counterpart for delivering mid-IR broad SC spectra, which opens new application gateways in optical coherence tomography and other mid-IR applications.

2. Proposed PQF Design

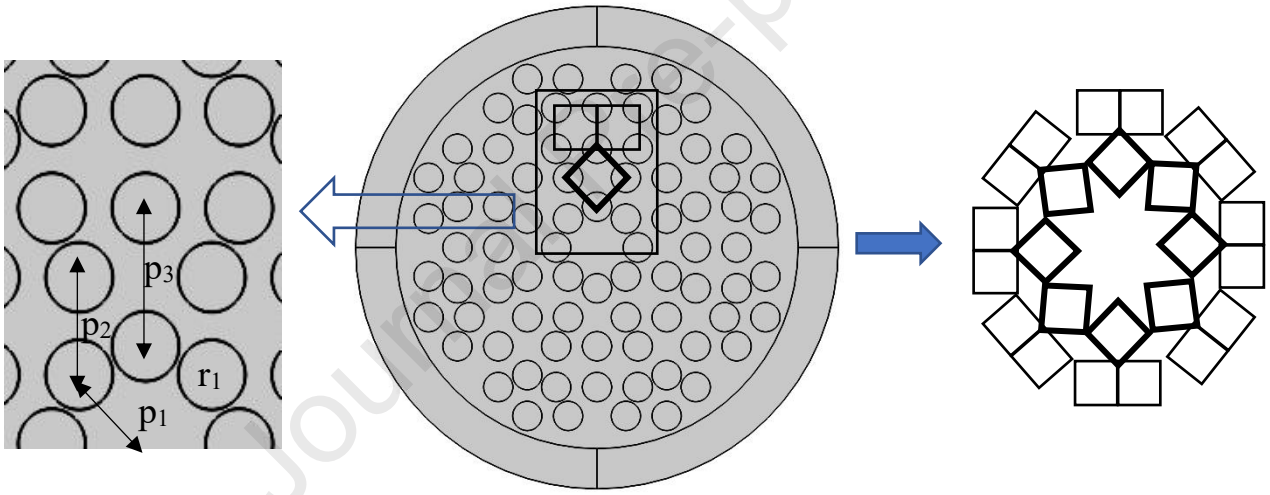


Fig. 1. Proposed Penrose-type PQF Structure

A Penrose-type PQF is designed with a circular core adjoining by cladding layers of air holes. The central core is specifically composed of Te-based $Ge_{20}Se_{70}Te_{10}$ chalcogenide glass. The radius of air holes is kept constant throughout the cladding of the proposed PQF structure, denoted by r_1. The respective air holes are located at two different lattice units, which satisfy eight-fold rotational symmetry without a translational symmetric nature. The two lattice units are one rhombus and one rectangle (divided into two squares) stacked over the other. We have defined three different types of pitch. p_1 is the first pitch parameter that defines the distance of the air hole of the first cladding layer from the centre of the PQF. Following p_1, p_2 and p_3 served as two distinct pitches which defines the centre-to-centre interspace of air holes of first and second layer of cladding and centre-to-centre stretch of air holes of first and third layer of cladding respectively. To eliminate the back reflection effect, a perfectly matched layer (PML) is applied as a layer surrounding the PQF structure. The Penrose-type PQF is proposed for the first time for SCG applications, which arises as a novelty factor for using this certain design.

The proposed PQF was successful in delivering a flat near zero dispersion and respective zero dispersion wavelength at 3.8 μm.

3. Linear and nonlinear characteristics

The GST based-$Ge_{20}Se_{70}Te_{10}$ is the chalcogenide glass implied in the PQF structure as the core material. The refractive index of the mentioned GST glass is defined as a wavelength-dependent Sellmeier equation, and the same is mentioned below with all the numerical parameters [39]

$$n^2(\lambda)=1+\frac{\left(5.225\times\lambda^2\right)}{\lambda^2\text{-}0.298^2}+\frac{\left(0.202\times\lambda^2\right)}{\left(\lambda^2\text{-}21.606^2\right)} \tag{1}$$

The spectral components of input optical pulse traverse at different phase velocities in PCF or PQF structures. The group velocity dispersion (GVD) parameter determines the limit at which the intervals between the phase velocities are created. The GVD parameter is totally dependent on the variation of the real part of the effective mode index.

$$D(\lambda)= -\frac{\lambda}{c}\frac{d^2n_{eff}}{d\lambda^2} \tag{2}$$

Equation (2) comprises the effect of both material and waveguide dispersion. The *β(ω)* parameter is also evolved with the dispersion calculation. The term *β(ω)* contains a discrete set of parameters obtained by expanding Taylor's series. Every term of *β(ω)* describes a characteristic of the optical wave. The effective mode index is well defined by the first term. The second term denotes the group velocity, and the third describes the GVD parameters. We have taken 9th-order consideration for the *β(ω)* parameter.

$$\beta(\omega) = \beta(\omega_o) + \beta_1(\omega_o)(\omega - \omega_o) + \frac{1}{2!}\beta_2(\omega_o)(\omega - \omega_o)^2 + \frac{1}{3!}\beta_2(\omega_o)(\omega - \omega_o)^2 \ldots \tag{3}$$

For generating broadband spectra, our ultimate aim should be the adoption of PCF parameters such that the magnitude of the effective mode area must exhibit a low magnitude such that it can successfully sustain modal fields at larger wavelengths and enhance the respective nonlinearity towards higher magnitude (in the order of 100 $W^{-1}km^{-1}$). Therefore, the effective mode area must be calculated using the following formula.

$$A_{eff} = \frac{\left(\iint_{-\infty}^{\infty}|E|^2dxdy\right)}{\left(\iint_{-\infty}^{\infty}|E|^4dxdy\right)} \tag{4}$$

The nonlinearity term comprises the effect of all the nonlinearities that participated in the fiber, like SPM, XPM, SRS, SBS, and many more. The overall quantification of nonlinearities is done by calculating the value of γ by the given formula.

$$\gamma = \frac{2\pi n_2}{\lambda A_{eff}} \tag{5}$$

Equation (5) shows the direct relationship of γ with the nonlinear refractive index and the inverse relationship with effective mode area and wavelength. The nonlinear refractive index of the proposed material $Ge_{20}Se_{70}Te_{10}$ is 8×10^{-18} $W^{-1}km^{-1}$ [39, 40].

4. Theoretical Model of Supercontinuum generation

For the simulation of SC in PQF, a generalized nonlinear Schrodinger equation (GNLSE) is formulated which comprises all the effects of linear and nonlinear characteristics of the proposed model. The GNLSE is solved using the split-step Fourier method (SSFM) and the same equation is described below [41].

$$\frac{dA}{dz}+\frac{\alpha}{2}A-\left(\sum_{n\geq 2}\beta_n\frac{i^{n+1}}{n!}\frac{\partial^n A}{\partial t^n}\right)=i\gamma\left(1+\frac{i}{\omega_o}\frac{\partial}{\partial t}\right)A(z,t)\int_{-\infty}^{\infty}R(t')\left|A(z,t-t')\right|^2dt' \quad (6)$$

The above equation denotes the optical field envelope as *A(z, t)* and fiber losses as *α*. The term *β(ω)* and its parameters are already discussed in the above section. *R(t)* represents the instantaneous functional form of the electronic and Raman contributions and the same can be written as [41]

$$R(t') = (1 - f_r)\delta(t) + f_r h_r(t') \quad (7)$$

In equation (7), fractional participation is denoted as fr=0.115 [42], which was responsible for the delayed response of Raman contributions to nonlinear polarization. $h_R(t)$ is represented as the Raman response function, which consists of the Raman period and lifetime denoted as τ_1 and τ_2, consists of values 23.1 fs and 195 fs, respectively [42], and the corresponding quantified function is described below [41]

$$h_r = \frac{\tau_1^2 + \tau_2^2}{\tau_1\tau_2^2}\exp\left(-\frac{t'}{\tau_2}\right)\sin\left(\frac{t'}{\tau_1}\right) \quad (8)$$

Relative coherence between the consecutive spectral evolutions of the SC is used to quantify changes in spectral phase over the Supercontinuum band. The following equation measures the first-order relative coherence [41]:

$$\left|g_{12}^{(1)}(\lambda, t_1 - t_2)\right| = \left|\frac{< E_1^*(\lambda, t_1)E_2(\lambda, t_1) >}{\sqrt{< |E_1(\lambda, t_1)|^2 >< |E_2(\lambda, t_2)|^2 >}}\right| \quad (9)$$

The degree of coherency of the produced SC spectra at optimal parameters is verified by the numerical magnitude for coherence, which should be closest to unity. The average coherency must be close to one to strongly agree with generated SC spectra.

5. Numerical Simulations and Observations

This section is divided into three segments which describe the dispersion characteristics of the proposed fiber and help in optimizing the parameters like r_1, p_1, p_2, p_3. A deep understanding of effective mode area and nonlinearity is also realized from this section. Numerical simulations for SC and corresponding observations are reported in the concluding segment.

5.1.Evaluation of Dispersion Characteristics

The normal and anomalous region's realization is obtained from the respective PQF's dispersion characteristics, which was displayed in fig. 2. The respective characteristics also give an understanding of the operating wavelength to be used for further SC simulations. Our eventual goal is to extract a flat-top near zero dispersion by optimization of the fiber parameters. For the sake of optimizing the parameters, we have kept all the pitch parameters like p_1, p_2, and p_3 constant at 6.6 μm, 6.7 μm, and 9.5 μm, respectively, and changed the magnitude of r_1 by

0.1 µm from 2.3 µm to 2.5 µm at the initial study. It is reported in fig. 2(a) that zero dispersion wavelength (ZDW) is shifted towards lower wavelength (blue-shifted) on the increment of r_1 by 0.1 µm. At 2.3 µm, the ZDW falls at 3870 nm, but when we increase the parameter r_1 by 0.1 µm, the ZDW begins decreasing by 50 nm. Therefore, it is concluded that influencing the magnitude of r_1 can vary the dispersion to a larger extent. In the second stage of the study, shown in fig. 2(b), we influenced the parameter p_1 by keeping constant magnitudes for r_1, p_2 and p_3 as 2.4 µm, 6.6 µm and 9.5 µm. The observation shows that on increasing the p_1 by 0.2 µm, from 6.4 µm to 6.8 µm, the ZDW is red-shifted from 3740 nm to 3910 nm. The same

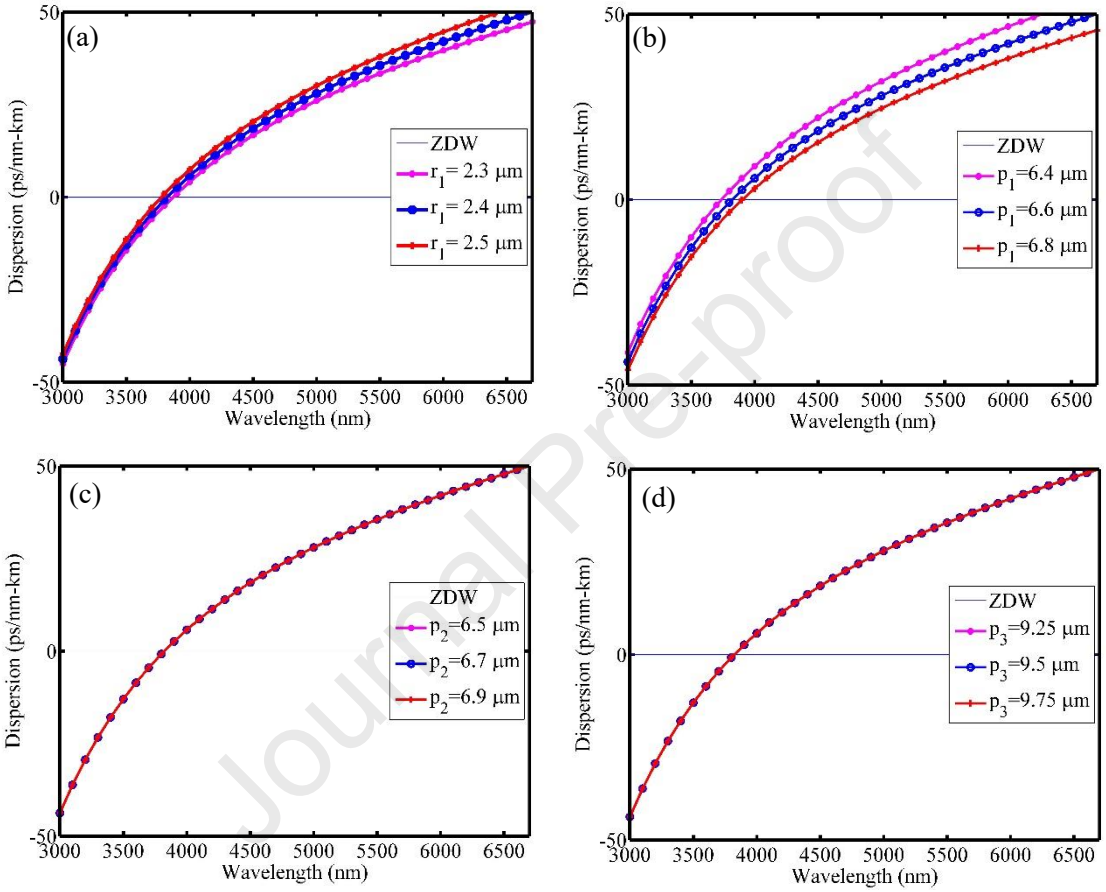


Fig. 2. Dispersion Characteristics on influencing (a) r_1 (b) p_1 (c) p_2 (d) p_3

statement concludes that affecting the p_1 can also influence the ZDW at a higher rate. Now, we set the parameters like r_1, p_1, and p_3 constant at 2.4 µm, 6.6 µm, and 9.5 µm and changed the measure of the second pitch parameter p_2 by 0.2 µm from 6.5 µm to 6.9 µm shown in fig. 2(c). The observation displays no change in ZDW on influencing the value of p_2. Similarly, we have extracted dispersion data by changing the value of p_3 from 9.25 µm to 9.75 µm at the span of 0.25 µm by keeping other parameters constant. The observation is identical to the former observation and offers no change in the magnitude of ZDW displayed in fig. 2(d). This signifies that p_2 and p_3 have no influence over dispersion. Therefore, by maintaining the balance between the best dispersion argument and fabrication possibility, we have optimized our parameters, and the same is tabulated in Table I. Furthermore, from the fabrication point of view, we have influenced the optimum parameters by ± 2% and extracted the dispersion manner illustrated in fig. 3. It is noticed that only ± 30 nm (~ 0.8 %) change (with respect to optimum ZDW, i.e., 3820 nm) is observed in ZDW respectively, which does not influence our ZDW highly. Our

resultant SC data will also be unaffected by this minor change. In the accordance to our calculated ZDW value, we have chosen our central wavelength as 4 μm, which lies in the anomalous region having a small magnitude of dispersion i.e., 6.5 ps/nm-km. A mode-locked thulium-based femtosecond laser operating at 4 μm can be used for this simulation [43]

Table I
Optimized PQF Parameters

	Radius of air holes (r_1)	1st Pitch (p_1)	2nd Pitch (p_2)	3rd Pitch (p_3)
PQF Parameters	2.4 μm	6.6 μm	6.7 μm	9.5 μm

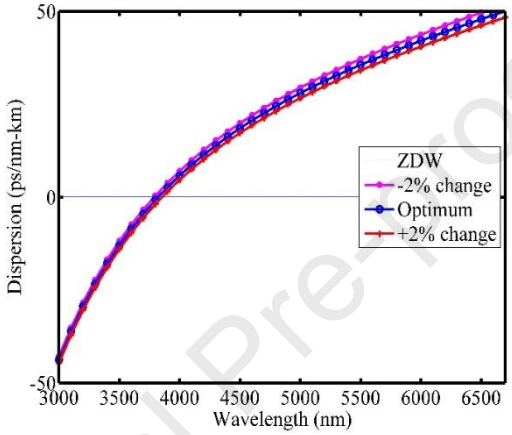


Fig. 3. Tolerance study of parameters at ±2% deviation

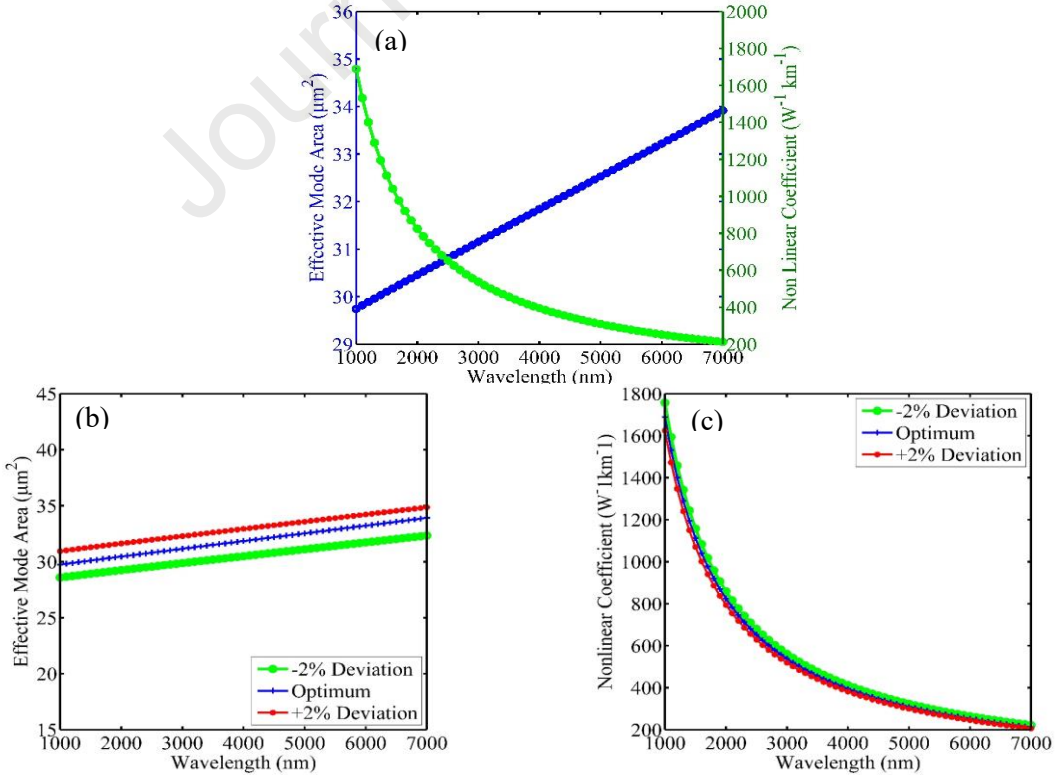


Fig. 4(a) Study of Effective Mode Area (EMA) and Nonlinearity at optimized parameters with the change of wavelength
(b, c) tolerance study of EMA and nonlinearity at ±2% deviations

Journal Pre-proof

5.2.Evaluation of Effective Mode Area and Nonlinearity

The estimation of effective mode area (EMA) and respective nonlinearity is displayed in Fig. 4(a). The figure concludes that EMA keeps increasing with the increment of wavelength, and the corresponding nonlinearity decreases as the wavelength increases. At 3.8 μm (ZDW at optimized parameters), the magnitude of EMA is 31.7 μm^2, and nonlinearity attains a larger value of 417 $W^{-1}km^{-1}$. The tolerance study is also done by variating the parameters by ± 2% and the same is displayed in fig 4(b, c). It is observed that ~3 % change is observed in the effective mode area (with respect to EMA at optimized parameters, i.e., 31.7 μm^2) and the respective nonlinearity observed the change of ~3% with respect to nonlinearity at optimum parameters. The same PQF may be fabricated by method of extrusion or drilling in future times [44, 45]. Compared to silica glass, the extrusion method allows for more flexibility in the creation of PCFs/PQFs derived from chalcogenide glasses and can withstand high dopant concentrations. This allows for easier modifications to the mechanical and optical properties of the material.

5.3.Evaluation of Supercontinuum Generation

For the generation of wide-band SC, a secant hyperbolic pulse is considered as the input pulse, and the corresponding pulse is expressed as:

$$U(z=0,t)=\sqrt{P_0}sech\frac{T}{T_0} \tag{10}$$

The numerical investigations were done by solving the GNLSE using the split-step Fourier method. The input pulse evolution in the PQF suffered a symmetric type broadening at the initial stage due to the participation of phase-matched conditions like SPM, dispersive effects, and other nonlinearities. At the later stage, the pulse broadening is the outcome of higher-order solitons generated in the PQF due to anomalous-region pumping. The appropriate action of the nonlinearities is clarified by calculated values of the distinct time scales tabulated in Table II, such as nonlinear length, dispersion length, soliton fission length, and order of solitons.

Table II
Length scales and other nonlinear parameters

Length Scale/ Parameter	Formula	Calculated Value
Peak Power	P_0	20 kW
Pulse Width	T_{FWHM}	50 fs
Pulse Duration	$T_0=\frac{T_{FWHM}}{1.763}$	28.3 fs
Group Velocity Dispersion	β_2	4.7359×10^{-2} ps^2/m
Nonlinear Length	$L_{NL}=\frac{1}{\gamma P_0}$	0.127 mm
Dispersion Length	$L_D=\frac{T_0^2}{\lvert\beta_2\rvert}$	16.91 mm

Soliton Order	$N = \sqrt{\frac{L_D}{L_{NL}}}$	≈12
Soliton Fission Length	$L_{fs} = \frac{L_D}{N}$	1.41 mm

From Table II, clear conclusions are drawn that after the propagation for 0.127 mm, nonlinear effects begin to act on the input pulse propagate over the proposed PQF, and dispersive effects get the participation at a fiber length of 16.91 mm. Since we are working on an anomalous region, the soliton fission length is also calculated, and the magnitude arrived at 1.41 mm, which gives a clear understanding of the position at which soliton fission will begin. The respective order of solitons is also calculated. Therefore, we begin our investigations by changing the fiber length beginning from 1 mm, and the same is displayed in Fig 5.

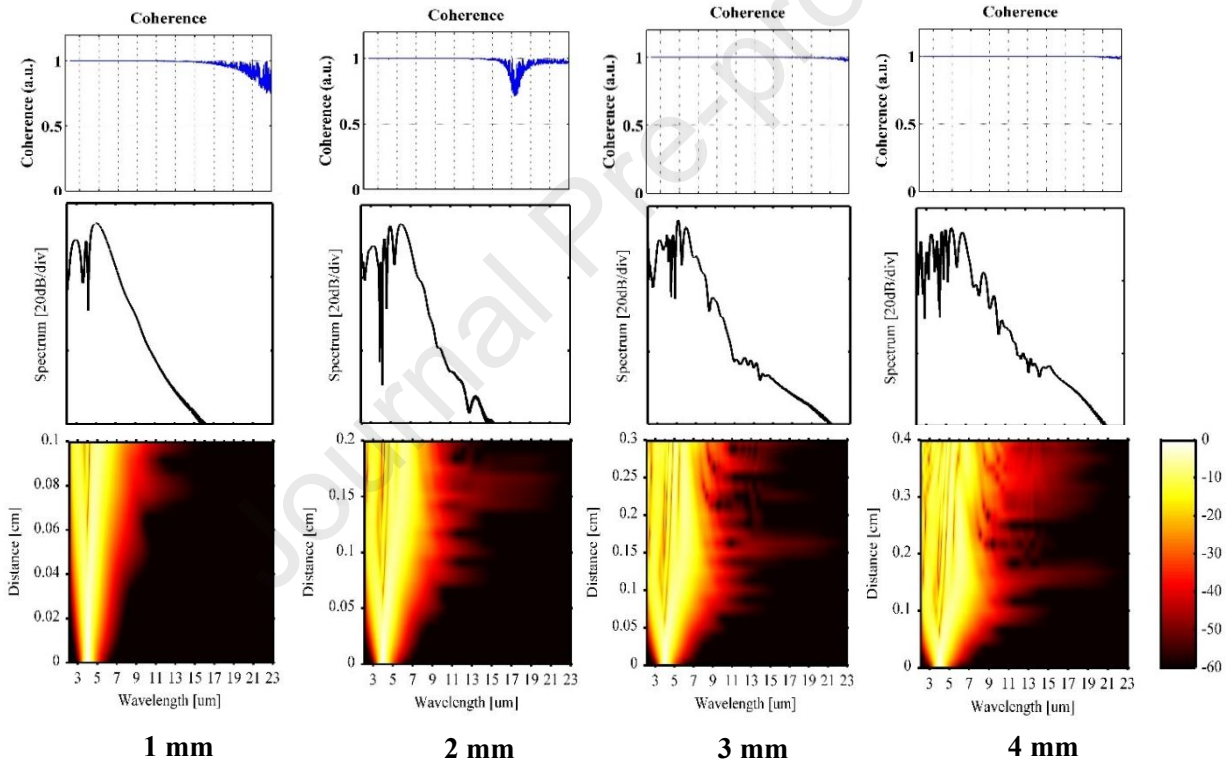


Fig.5. Numerically Simulated SC spectra at different fiber lengths at 4 μm pumping
(Top-Bottom) Coherence, Line Plot, Spectral Plot

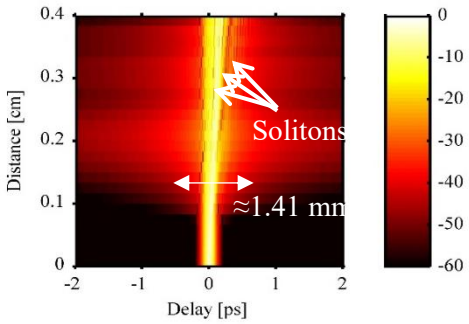


Fig. 6. Generation of solitons and soliton fission after traversing soliton fission length

It has been noticed that the SC spectra are broadened with the increment of fiber length. Till 4 mm, it kept on broadening with a highly coherent nature. Since 4 mm is greater than our calculated nonlinear length, nonlinear effects have started to act and ultimately broaden the band. Self-phase modulation is the reason behind such widening, which is due to the domination of SPM over dispersive effects, as the fiber length is smaller than the calculated dispersion length.

After propagating 1.41 mm (soliton fission length), the higher-order solitons started to break down into fundamental solitons displayed in fig. 6. The corresponding solitons suffered from a red shift due to the action of Raman-self frequency transfer. At the final stage of all the nonlinear actions, our SC spectrum is generated having the range of 2-21 μm (at 60 dB power level) devised with a highly coherent nature.

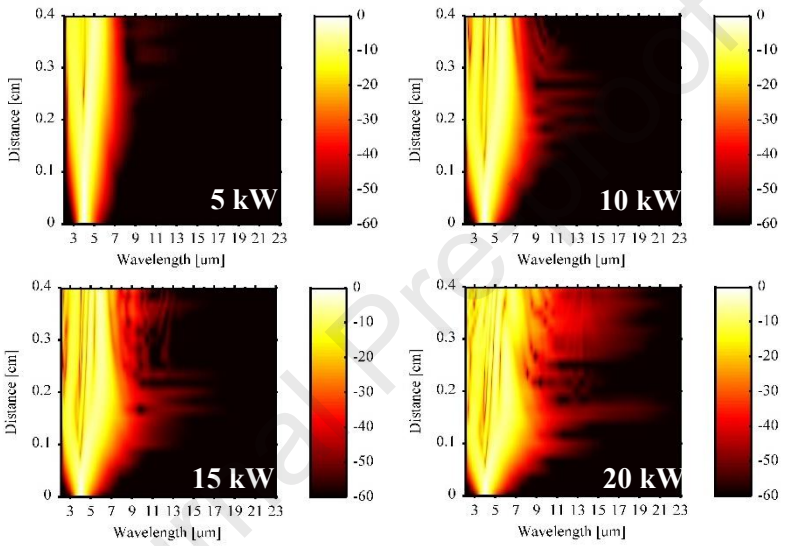


Fig. 7. Numerically Simulated SC spectra at different peak powers at 4 mm length of fiber

In the following study, we have influenced the peak power of the input pulse with a constant fiber length of 4 mm and pulse width of 50 fs. The results are exhibited in Fig. 7. Since nonlinear length is inversely proportional to the peak power, the subsequent increment in peak power results in a decrement in the nonlinear length and correspondingly increases the bandwidth of SC in a very short fiber length. Therefore, on increasing the peak power up to 20kW, the spectra keep on broadening. However, after 20 kW, the resulting spectra turn out to be noisy and disrupt the coherency. Therefore, we have taken peak power 20kW as our ideal operating peak power. Since dispersion length is directly proportional to pulse width,

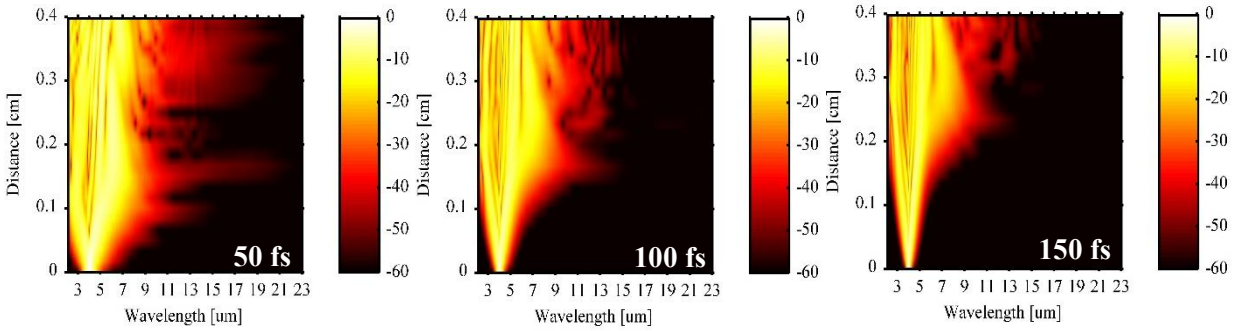


Fig.8. Numerically Simulated SC spectra at different pulse width at 4 mm length of fiber with peak power of 20 kW

escalating the magnitude of pulse width automatically increases the dispersive length, which ceases the wide-band SC at a smaller fiber length shown in fig. 8. Nonlinear length is independent of pulse width. The time-delayed components are clearly observable at FROG

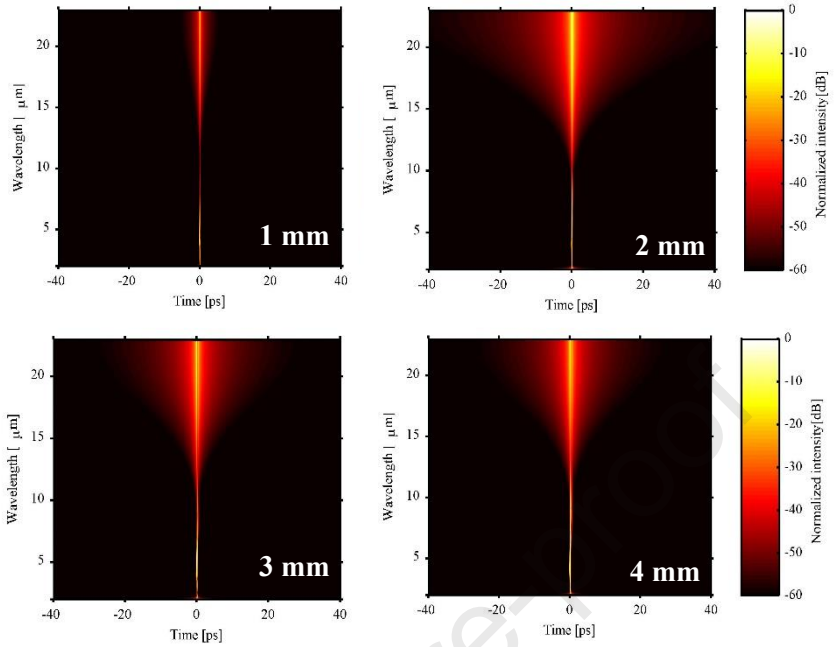


Fig. 9. Time-Wavelength FROG spectrogram

spectrograms manifested in fig. 9. The information about pulse shape, variation in optical pulse, and frequency chirp can be gathered from FROG spectrograms.

After performing all the studies, the conclusion can be drawn that at the input of femtosecond laser having a wavelength of 4 μm, a wide band spectra of 2-21 μm is extracted at 20kW peak power and 50fs pulse width in the fiber length of 4 mm. Due to the small fiber length and moderate peak power, the proposed Penrose-based PQF can be a perfect candidate for performing SCG and a better counterpart for applications like optical metrology (OM), OCT, etc.

Table III
Comparison of extracted results with previously done works

Sr. No.	Material & Fiber type	SCG spectra	Broadening	Peakpower & Fiber Length	Ref.
1.	As_2Se_3 based PCF	2-15 μm	13 μm	3 kW & 5 mm	[47]
2.	As_2Se_3 based PCF	2.53-16.6 μm	14 μm	4 kW & 47 mm	[48]
3.	As_2S_5 based PCF	0.7-5.2 μm	4.5 μm	25 kW & 5 mm	[49]
4.	$Ge_{20}Se_{70}Te_{10}$ based PCF	1.7-18 μm	16.3 μm	200 kW & 40 mm	[38]
5.	$Ge_{20}Se_{70}Te_{10}$ based PCF	3.25-13 μm	9.75 μm	190 kW & 40 mm	[39]

6.	$Ge_{11.5}As_{24}Se_{64.5}$ based PQF	2-15 μm	13 μm	2 kW & 8 mm	[33]
7.	As_2Se_3 based PQF	1-10.2 μm	9.2 μm	6 kW &15 mm	[34]
8.	As_2Se_3 based PQF	1.2-9.3 μm	8.1 μm	10 kW & 5 mm	[46]
9.	*$Ge_{20}Se_{70}Te_{10}$ based PQF This Work*	*2-21 μm*	*19 μm*	*20 kW & 4 mm*	

6. Conclusion

In this proposed work, we have designed a Penrose-based PQF containing air holes of the same magnitude. The distribution of air holes does not inherit any translational symmetry among them but exhibits eight-fold rotational symmetry. We have calculated the dispersion characteristics of our proposed PQF, which delivers a flat-top dispersion model inheriting ZDW at 3.8 μm. Furthermore, we have also calculated the effective mode area having a magnitude of 31.7 μm^2 and respective nonlinearity having a magnitude of 417 $W^{-1}km^{-1}$ at extracted ZDW. We have adopted 4 μm as our pump wavelength, which falls in the anomalous region, and successfully demonstrated wide-band spectra of 2-21 μm at a peak power of 20kW and pulse width of 50 fs, which wraps the transparency range of the proposed material and Mid-IR range very efficiently. The same observation is generated at the very small length of fiber, i.e., 4 mm, which is the key highlight of our work. Apart from these demonstrations, we have also done the tolerance study by changing the measurement of parameters by ±2%, which is a very important aspect from the fabrication point of view. The generated wide-band SC opens the doors for the Penrose-based PQF for the Mid-IR applications such as optical metrology, optical coherence tomography, etc., and can be exploited to become a perfect candidate for nonlinear applications.

7. Acknowledgement

Akash Khamaru highly acknowledges the support from the Delhi Technological University, Delhi, for providing the research fellowship and necessary support to conduct this research work.

8. References

[1] Reintjes, J. F. (2020). Stimulated Raman and Brillouin scattering. In *CRC Handbook of Laser Science and Technology Supplement 2* (pp. 334-364). CRC Press.

[2] C. Khallouf, V. T. Hoang, G. Fanjoux, B. Little, S. T. Chu, D. J. Moss, & T. Sylvestre, Raman scattering and supercontinuum generation in high-index doped silica chip waveguides. In Nonlinear Optics and its Applications 2024, Vol. 13004, pp. 73-81, (2024) SPIE.

[3] A. V. Husakou, & J. Herrmann, Supercontinuum generation, four-wave mixing, and fission of higher-order solitons in photonic-crystal fibers. JOSA B, 19(9), (2002) 2171-2182.

[4] R. R. Alfano, Y. R. Shen, G. Z. Yang, Theory of self-phase modulation and spectral broadening. The Supercontinuum Laser Source: The Ultimate White Light, (2016) 1-32.

[5] H. Sone, T. Arai, M. Imai, Y. Imai, Modal birefringence dependent supercontinuum due to cross-phase modulation in a dispersion-flattened/decreasing fiber. Optical review, 9, (2002) 89-93.

[6] G.P. Agrawal, & N. A. Olsson, (1989). Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers. IEEE Journal of Quantum Electronics, 25(11), (1989) 2297-2306.

[7] G.P. Agrawal, Nonlinear Fiber Optics, 5th ed., Elsevier Academic Press, 2013.

[8] P. Russell, Photonic crystal fibers, Science 299 (2003) 358–362.

[9] J.M. Dudley, G. Genty, S. Coen, Supercontinuum generation in photonic crystal fiber, Rev. Mod. Phys. 78 (2006) 1135–1184.

[10] R.K. Sinha, S.K. Varshney, Dispersion properties in photonic crystal fibers, Microw. Opt. Tech. Lett. 37 (2003) 129–132.

[11] B. Dabas, R.K. Sinha, Dispersion characteristics of hexagonal and square lattice chalcogenide As_2Se_3 glass photonic crystal fiber, Opt. Commun. 283 (2010) 1331–1337.

[12] B. Dabas, R.K. Sinha, Design of highly birefringent chalcogenide glass PCF: a simplest design, Opt. Commun. 284 (2011) 1186–1191.

[13] R. Cherif, "Highly nonlinear As_2Se_3-based chalcogenide photonic crystal fiber for midinfrared supercontinuum generation," Opt. Eng. 49 (2010), 095002.

[14] D. S. Tomer, A. Kumar, A Graded Index Hybrid Photonic Crystal Fiber for Supercontinuum Generation using $AsSe_2$-As_2S_5. Journal of Optics, (2024) 1-10.

[15]. A. Bétourné, A. Kudlinski, G. Bouwmans, O. Vanvincq, A. Mussot, and Y. Quiquempois, "Control of supercontinuum generation and soliton self-frequency shift in solid-core photonic bandgap fibers," Opt. Lett. 34 (2009) 3083–3085.

[16] W. Su, S. Lou, H. Zou, B. Han, Design of a highly nonlinear twin bow-tie polymer photonic quasi-crystal fiber with high birefringence. Infrared Physics & Technology, 63, (2014) 62-68.

[17] E. Liu, W. Tan, B. Yan, J. Xie, R. Ge, R., & Liu, J. (2018). Broadband ultra-flattened dispersion, ultra-low confinement loss and large effective mode area in an octagonal photonic quasi-crystal fiber. JOSA A, 35(3), 431-436.

[18] B. Ung and M. Skorobogatiy, "Chalcogenide microporous fibers for linear and nonlinear applications in the mid-infrared," Opt. Express 18 (2010) 8647–8659.

[19] R.R. Alfano, S.L. Shapiro, Emission in the region 4000 to 7000 A via four-photon coupling in glass, Phys. Rev. Lett. 24 (1970) 584.

[20] J.K. Ranka, R.S. Windeler, A.J. Stentz, Visible continuum generation in air silica microstructured optical fibers with anomalous dispersion at 800 nm, Opt. Lett. 25 (2000) 25–27.

[21] T.S. Saini, A. Kumar, & R. K. Sinha, Broadband mid-infrared supercontinuum spectra spanning 2–15 μm using As_2Se_3 chalcogenide glass triangular-core graded-index photonic crystal fiber. Journal of lightwave technology, *33*(18) (2015), 3914-3920.

[22] T. S. Saini, A. Kumar, & R. K. Sinha, Broadband mid-IR supercontinuum generation in As_2Se_3 based chalcogenide photonic crystal fiber: A new design and analysis. Optics Communications, *347* (2015), 13-19.

[23] A. Medjouri and D. Abed, Design and modelling of all-normal dispersion $As_{39}Se_{61}$ chalcogenide photonic crystal fiber for flat-top coherent mid-infrared supercontinuum generation, Opt. Fiber Technol., **50** (2019) 154–164.

[24] A. Medjouri and D. Abed, Mid-infrared broadband ultraflat-top supercontinuum generation in dispersion engineered Ge-Sb-Se chalcogenide photonic crystal fiber, Opt Mater *(Amst)*, **97**, (2019)

[25] A. Khamaru, A. Kumar, As_{38} Se_{62} based segmented clad-graded index photonic crystal fiber for supercontinuum generation covering 3–9.5 μm with moderate peak power. Opt Quant Electron **56**, 1246 (2024).

[26] Wen, Jin, et al. "Coherence analysis of supercontinuum generation in nitrobenzene liquid-core photonic crystal fiber based on adaptive step-size methods." Opt. Quant Electron 56.4 (2024), 1-17.

[27] N. Thi, H.T. Duc, and C. V. Lanh, "Optimization of dispersions in benzene-core square photonic crystal fibers for mid-infrared supercontinuum generation with very low peak power." Optik 304 (2024), 171754.

[28] S. Kim, C. Kee, J. Lee, Novel optical properties of six-fold symmetric photonic quasicrystal fibers, Opt. Express 15 (2007) 13221-13226.

[29] E. Liu et al. Guiding characteristics of sunflower-type fiber. Superlattices and Microstructures, *115*, (2018), 123-129.

[30] S. Sivabalan, J. Raina, High normal dispersion and large mode area photonic quasi-crystal fiber stretcher, IEEE Photonics Technol. Lett. 23 (2011) 1139-1141.

[31] E. Liu, & J. Liu, J. Quasiperiodic photonic crystal fiber. Chinese Optics Letters, *21*(6), (2023), 060603.

[32] R. Ahmad, M. Komanec, & S. Zvanovec Circular lattice photonic crystal fiber for mid-IR supercontinuum generation. IEEE Photonics Technology Letters, *28*(23) (2016), 2736-2739.

[33] M. Valliammai, & S. Sivabalan, Wide-band supercontinuum generation in mid-IR using polarization maintaining chalcogenide photonic quasi-crystal fiber. Applied Optics, *56*(16), (2017), 4797-4806.

[34]T. Zhao et al. Highly-nonlinear polarization-maintaining As_2Se_3-based photonic quasi-crystal fiber for supercontinuum generation. Optical Materials, *73*, (2017), 343-349.

[35] V. Shiryaev, M. Churbanov, Trends and prospects for development of chalcogenide fibers for mid-infrared transmission, J. Non-Cryst. Solids 377 (2013) 225–230

[36] K. T. Kitchin, Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. Toxicology and applied pharmacology, *172*(3), (2001), 249-261.

[37] A. Lemiere et al., Mid-infrared supercontinuum generation from 2 to 14 μm in arsenic- and antimony-free chalcogenide glass fibers. JOSA B, *36*(2), (2019), A183-A192.

[38] A. Lemière et al., 1.7–18 μm mid-infrared supercontinuum generation in a dispersion-engineered step-index chalcogenide fiber. Results in Physics, *26*, (2021), 104397.

[39] A. Lemière et al., Mid-infrared two-octave spanning supercontinuum generation in a Ge–Se–Te glass suspended core fiber. Laser Physics Letters, *16*(7), (2019), 075402.

[40] L Sun et al. (2016). Investigation of the third-order nonlinear property of Ge–Se–Te glasses at mid-infrared. Applied Physics A, *122*, 1-6 (2016).

[41] G. P. Agrawal, Nonlinear fiber optics. In *Nonlinear Science at the Dawn of the 21st Century* (pp. 195-211). Berlin, Heidelberg: Springer Berlin Heidelberg (2000).

[42] W. Gao et al., Simulation and analysis of supercontinuum generation in the waveband up to 25 μm. Applied Optics 61, (2022) 6697-6703.

[43] N. Leindecker, A. Marandi, R. L. Byer, K. L. Vodopyanov, J. Jiang, I. Hartl, M. Fermann, and P. Schunemann, "Octave-spanning ultrafast OPO with 2.6–6.1 μm instantaneous bandwidth pumped by femtosecond Tm fiber laser," Opt. Exp., vol. 20, (2012) pp. 7046–7053,

[44] B. J. Eggleton, B. Luther-Davies, and K. Richardson, "Chalcogenide photonics," Nat. Photonics 5, (2011) 141–148.

[45] H. E. Heidepriem and T. M. Monro, "Extrusion of complex preforms for microstructured optical fibers," Opt. Express 15, 15086–15092 (2007).

[46] R. Ahmad, M. Komanec, & S. Zvanovec, Circular lattice photonic crystal fiber for mid-IR supercontinuum generation. IEEE Photonics Technology Letters, 28(23), (2016) 2736-2739

[47] Q. Peng et al., "Single-Mode Segmented Cladding Chalcogenide Glass Fiber With Ultra-Large Mode Area," Journal of Lightwave Technology, vol. 41, no. 17, pp. 5722–5728, 2023.

[48] X. Feng et al., "Ultra-Large Mode Area Chalcogenide Photonic Crystal Fiber for High-Power Mid-Infrared Applications," in 2020 Conference on Lasers and Electro-Optics (CLEO), 2020, pp. 1–2.

[49] M. Diouf, A. Wague, and M. Zghal, "Numerical investigation of an ultra-broadband coherent mid-infrared supercontinuum in a chalcogenide $AsSe_2$-As_2S_5 multimaterial photonic crystal fiber," J. Opt. Soc. Am. B, vol. 36, no. 2, pp. A8–A14, Feb. 2019.

Journal Pre-proof

Highlights

- The Penrose-based Photonic Quasi Crystal Fiber is proposed for the first time for Supercontinuum (SC) Generation.
- The GST based-$Ge_{20}Se_{70}Te_{10}$, a non-toxic, Arsenic-free highly nonlinear chalcogenide glass is used for the PCF composition.
- A wide band of SC (2-21 μm) is generated at moderate peak power (~20 kW), which is very less as compared to previous Te-based chalcogenide PCF which had utilized 200 kW of peak power.
- The broadband extracted SC is able to wrap the proposed material's transparency range (2-25 μm) and also covers the Mid-IR range simultaneously.

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575336

Hate Speech Detection using CoT and Post-hoc Explanation through Instruction-based Fine Tuning in Large Language Models

Palak Gupta
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
palakgupt794@gmail.com

Nikhilesh Jain
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
nikhilesh.jain07@gmail.com

Aruna Bhat
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
aruna.bhat@dtu.ac.in

***Abstract*—In recent times, both social media platforms and researchers have been working to identify hateful language using large language models. However, none of these efforts have focused on using instruction-based fine-tuning with explanations and additional context information in the detection process. This research study has conducted experiments with LaMini and Google Flan-T5 models to analyze the impact of CoT prompts and explanations in zero shot and few shot settings to enhance model interpretability. The analysis reveals that on average including target information in pipeline significantly improves model performance. There is also a considerable effect of adding the explanations into the pipeline, boosting accuracy. The research findings highlight the importance of context and structured prompts in improving model accuracy, highlighting the need for responsible development and deployment of hate speech detection systems in online environments.**

***Index Terms*—Hate Speech Detection, Contextual Information, Large Language Models, Lamini Model, Google Flan-T5, Prompts**

I. Introduction

The prevalence of hate speech on modern online social media platforms has become a significant concern. More and more individuals are falling victim to online harassment, abuse, and cyberbullying, causing substantial trauma and psychological distress [1] Language has a powerful ability to spread social prejudices and stereotypes, which are frequently hidden under layers of implicit meaning rather than being explicitly stated [2] [3]. To address this troubling trend, social media platforms have implemented moderation policies to limit the spread of hate speech and abusive content. A crucial aspect of content moderation involves identifying and filtering out abusive language.

Typically, this is done by training language models on human-annotated data for classification. However, this approach comes with challenges, including the significant resources required in terms of labour and expertise to annotate such hateful content. Moreover, the process of annotation exposes individuals to disturbing material, which can take a toll on their psychological well-being. As a result, recent efforts have explored the potential of Large Language Models (LLMs) for detecting hate speech [4]. Interestingly, studies on intersection of hate speech detection and LLMs have not yet deeply examined the role of additional contextual information as input or output to LLMs. For LLM systems, attempting to negotiate the complex web of social relationships, the identification and understanding of hate biases are critical needs because a lack of these can lead to the deployment of harmful technologies [3]. This suggests a potential avenue for further research in enhancing the effectiveness of LLMs in combating hate speech online.

This research focuses on hate speech detection—an effort to identify (online) language that manifests biases and stereotypes in harmful ways. This task commonly requires generating a comprehensive full-text explanation alongside a classification for human content moderators to review.

Remarkably, this explanation-seeking is consistent with current advances in the field of LLMs [5] [6]. Typically, task-specific instructions and/or demonstration examples are used to urge these models during their evaluation in the zero-shot or few-shot scenario. The few-shot prompting technique takes advantage of in-context learning, in which models ”learn” to improve their decision-making in response to the context that they are given. Although this approach has shown results, a current research line along similar lines suggests that adding explanations to the prompting process can further improve these models' in-context performance [7].

Explanations give information structure and logic, which helps the models analyse data more logically and enable better decision-making. Pre-answer chain-of-thought (CoT) reasoning and post-answer explanations are the two main approaches to integrating explanations. The former has demonstrated potential in challenging tasks requiring iterative reasoning, such as arithmetic and logical reasoning [8]. It generates intermediate reasoning steps prior to the final predictions. On the other hand, the latter produces answers ahead of the explanations that go along with them. This has been shown to improve few-shot learning in LLMs by creating a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:37:23 UTC from IEEE Xplore. Restrictions apply.

more abstract reasoning process, which positively alters task inference [5].

Considering this, we pose the following research question: Can we prompt LLMs to reason about a potentially hateful post and subsequently produce a prediction or vice versa? In light of this, we explore whether an inductive (i.e. post-answer explanations) or deductive (i.e. CoT) approach leads to better performance on hate speech datasets with full-text annotations. To answer these questions, we use the SocialBiasFrames dataset **[3]** as our primary data source. We include both distilled versions of larger models and a full-size non-distilled model in our analysis.

II. Related Work

A. Explainable Hate Speech Detection

Hate speech detection (HSD) has attracted significant attention in recent years, with early research primarily focusing on binary classification of hateful, abusive, or toxic language [9] [10]. However, the importance of not only classifying but also understanding the model's reasoning behind HSD has become increasingly relevant, as it could enable automatic flagging or AI-augmented writing interfaces to analyse potentially harmful content with detailed explanations for users or moderators to verify [3].

The HateExplain dataset, presented by [11] includes span annotations at the word and phrase levels that record human reasoning behind labelling. They noticed that models with high performance metrics and bias scores, such BERT [12], have trouble with measures related to plausibility and explainability. Using the SocialBiasFrames dataset, [3] trained models based on OpenAI's GPT and GPT2 [13] [14] to produce implicit power dynamics in textual form and categorize a post's offensiveness. These models were effective in categorizing speech patterns that conveyed unwanted social bias, but they struggled to provide comprehensive explanations.

The LatentHatred dataset [4] [15] used ChatGPT to categorize samples and produce explanations that matched. With an accuracy of 80%, ChatGPT was highly likely to be accepted by laypeople, even in circumstances where there was 20% disagreement. The algorithm produced high-quality annotations for implicit hate speech that were on par with those provided by human annotators. This allowed users to verify the implicit hatefulness of tweets more quickly with the use of more detailed visuals. However, [16] hypothesized that the prompting approach may have a significant impact on ChatGPT's performance on HSD tasks.

B. Instruction Based Fine-Tuning

A variety of methods for training and refining models have been studied in the past [17] [18]. These methods include starting from scratch and fine-tuning them to provide intermediate reasoning stages. The use of language models to predict intermediate computational results line-by-line for Python programs was first introduced by [19], outperforming direct final output prediction in terms of performance. Following the emergence of many-shot prompting strategies as recommended in [20], [5] demonstrated the effectiveness of including explanations in addition to examples in a few-shot prompt, improving language model performance on Big-Bench assessment tasks. Notably, they highlighted the effectiveness of explanations tuned using a validation set, showcasing improvements even with untuned explanations that outperformed meticulously matched control conditions. [21] critically evaluate the cross-domain performance of fine-tuned LLMs for hate speech detection, emphasizing the importance of model fine-tuning for nuanced understanding. In another study, [22] explore zero-shot learning with LLMs for detecting hate speech, showcasing the potential of prompt-based approaches in low-resource settings.

By using few-shot chain-of-thought (CoT) prompting, [23] investigated the natural growth of reasoning abilities inside suitably large language models. Results from their research showed significant improvements in arithmetic, commonsense, and symbolic reasoning performance. Furthermore, the effectiveness of CoT prompting was examined by [7] within the framework of the Big-Bench-Hard assessment suite, which is a subset of especially difficult tasks where model performances frequently fall short of human evaluations. As seen by the Codex model code-davinci-002 [24] outperforming average human-rater performance on 17 out of 23 evaluation tasks, their results demonstrated the superiority of CoT prompting. Furthermore, these developments highlight how important instruction-based fine-tuning approaches are for enhancing the logic and performance of models.

C. Models

LaMini [25]is a set of compact, effective language models, distilled from ChatGPT and trained on a sizable dataset consisting of 2.58 million instructions. Instructions from previous datasets, including self-instruct [26], P3 [27], FLAN [28], and Alpaca are compiled into this dataset. [25] employed ChatGPT (GPT-3.5-turbo) to produce additional instructions in order to augment the dataset. The focus was on diversifying these instructions while maintaining existing human-written instructions in the prompt. After that, GPT-3.5-turbo was used to produce answers for every instruction. Following the creation of the dataset, the authors improved a number of smaller language models on the instruction-response pairs, ranging in size from 61 million to 1.5 billion, and using different architectures (encoder-decoder and decoder-decoder). The T5 [29], Flan-T5 [30], CerebrasGPT [31], GPT-2 [14], and GPT-Neo [32] sources were the five sources used to initialize these models. The proposed models achieved comparable performance to Alpaca-7B while being significantly smaller, ranging from 4.5 to 10 times fewer parameters.

III. Methodology

A. Pre-processing

For our analysis, we utilise the test split of the SocialBiasFrames dataset, originally comprising 12,578 entries. . All columns other than "post" and "offensiveYN," which contain the textual post and the associated offensiveness ratings as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:37:23 UTC from IEEE Xplore. Restrictions apply.

annotated by a human, are removed during the first pre-processing step. We then manually aggregate posts and remove examples with empty labels. This phase involves averaging the offensiveYN scores (which range from 0 to 1) that each human annotator assigned to the post. A score of 0.5 or above indicates that the post is offensive; otherwise, it is classified as non-offensive and indicated with a ”no.” The posts are then made uniform by changing all Unicode characters to ASCII. In the end, 4691 posts are found, of which 2710 are marked as ”yes” and 1981 as ”no.”

B. Experimental Setup

Utilizing the LaMini models [25] on the resulting subset of the SocialBiasFrames dataset, we ran several tests. We also incorporated a non-distilled model—specifically, Google’s Flan-T5 variant—into our research so that our findings can be validated more broadly [30].

Four different experiments are included in our study to evaluate the performance of the models. First, we used a simple bare-bones zero-shot prompt to set the starting point for each model’s performance on the HSD task. Secondly, we investigated if using a zero-shot CoT prompt enhances the models’ performance. Thirdly, we used a few-shot prompt with post-answer explanations to examine if the models could learn from several examples. Lastly, by reversing the explanation and prediction’s sequence, we investigated the differences between the deductive and inductive approaches. We only used 4 shots for the latter two prompts because any more would have filled some models’ context windows.

The perplexity score, a gauge of how ”surprised” a model is to receive a particular input, affects how successful a prompt is [33]. We developed a “seed prompt” for the HSD task in accordance with [33] which was utilized to prompt the gpt-3.5-turbo model to produce potential variants. The prompt with the lowest perplexity was chosen as the first ”Question” in our prompts after the perplexities of each version and the seeds themselves were computed. We used beam search with a beam size of three for decoding. In order to prevent repetitive outputs, we also disabled early halting and made sure the output didn’t contain the same bigram twice. Owing to computational limitations, the explanation that the model produced could only contain a maximum of 100 tokens. The test was conducted through multiple iterations to train the model on context with the assistance of prompts, including instruction-based fine-tuning. The total computation time for all models exceeded 200 GPU hours.

C. Instruction Tuning through Prompts

The prompts we employed for each of the four strategies are displayed below in **Fig.1 - 4**. The post to evaluate is used in place of {post} at runtime.

D. Evaluation

The advantages of post-answer explanations were emphasized by [5] since they do not interfere with the evaluation pipeline. This is so that we may save computational resources

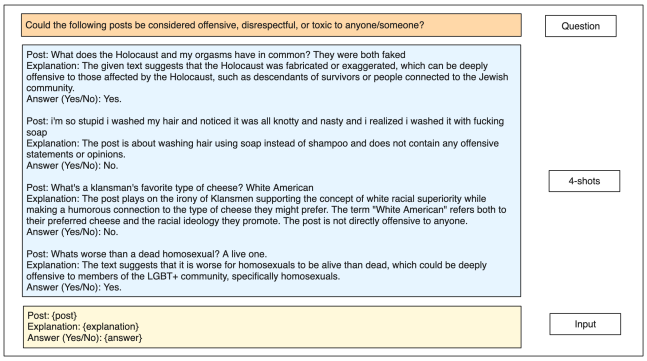


Fig. 1. default_4shot_cot_expl_first Prompt

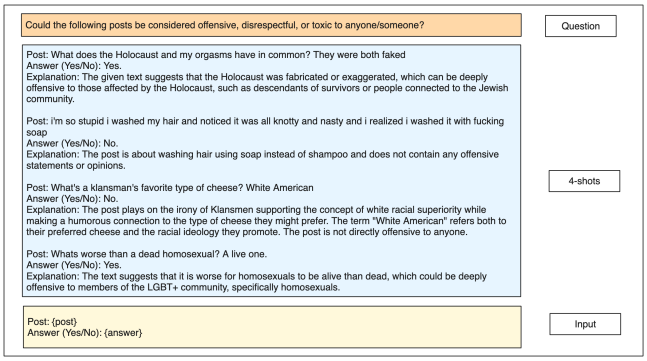


Fig. 2. default_4shot_cot_ans_first Prompt

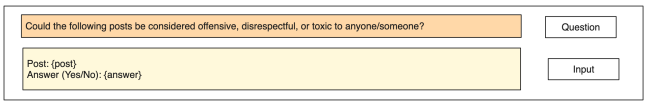


Fig. 3. default_0shot_no_cot Prompt

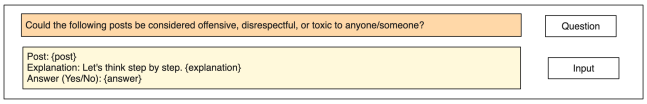


Fig. 4. default_0shot_cot_expl_first Prompt

by stopping computation after the first token is generated. In contrast, pre-answer reasoning chains need the assessment function to extract the solution from the model’s output, which adds complexity to the evaluation process. We use the two-step prompting approach suggested by [8] to handle this. We first entered the prompts into the model for each chain-of-thought prompting approach up until the {explanation} marker. Then, to produce a ”yes” or ”no” prediction, we replace {explanation} with the entire generated content of the model and enter the prompt up to {answer} again into the model. We basically employ the “reasoning” prompt as a tool to extract a comprehensive explanation from a language model. Subsequently, the “answer” prompt is utilised to extract the final response as shown in Fig. 5 Prompt pipeline. Using a SoftMax function, the output logits for the positive and negative classes were normalized and compared during the evaluation process. The language tokens for the positive class were [”Yes,” ”yes,” and [”No,” ”no”]. The tokens for the negative class were [”No,” ”no.” The label ”yes” was applied if the likelihood of the affirmative class was more than 0.5; if

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:37:23 UTC from IEEE Xplore. Restrictions apply.

not, the label ”no” was applied.

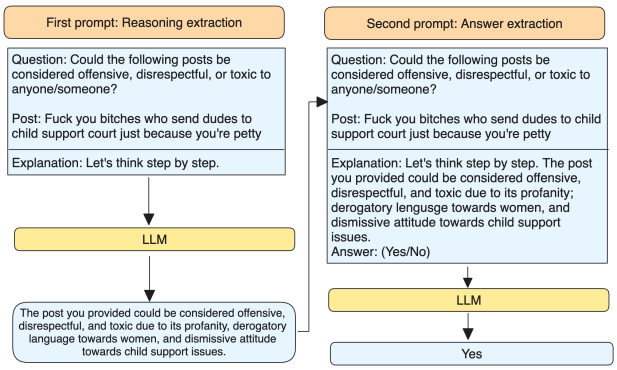


Fig. 5. Prompt Pipeline

By undergoing multiple runs on the LLM, we ensured a rich contextual understanding, crucial for accurate response generation. Furthermore, prompts served as navigational cues, aiding the model in focusing on relevant details and improving hate speech detection. This iterative refinement of prompts significantly boosted the model's performance, ensuring a deeper comprehension of the input data and improving its ability to identify and classify instances of hate speech accurately. Moreover, we calculated the average of the top three accuracies from these runs to conduct a thorough analysis of the model's effectiveness in hate speech detection tasks.

IV. Result and Analysis

Tables I - VI present the performance metrics of various LaMini and Google Flan-T5 models on the Hate Speech Detection (HSD) task across different prompting strategies and shot settings.

Starting with the 0-shot baseline, we observe that most models demonstrate improved performance compared to a ”constant” model that always outputs the same class. Notably, LaMini-Cerebras-256M and LaMini-Neo-125M consistently exhibit high accuracies across all prompting strategies, with a increase of 16.1% and 6.45%, respectively, when utilizing CoT compared to the baseline. Google Flan-T5-base model witnesses a notable 8.6% accuracy increase when incorporating CoT, highlighting the impact of contextual prompts in improving model performance.

Transitioning to the 4-shot settings, LaMini-Cerebras-256M and LaMini-Neo-125M continue to demonstrate strong performance, with both models experiencing moderate accuracy improvements of 24.1% and 22.5%, respectively, when employing the explain-first prompting strategy. LaMini-Flan-T5-248M and LaMini-T5-223M models display substantial accuracy boosts of 37.03% and 42.3%, respectively, with the post-answer explanation strategy, underscoring the efficacy of this approach in enhancing hate speech detection capabilities.

Moreover, the graphical representation of these results in Fig. 12. visually emphasizes the impact of different prompting strategies on hate speech detection accuracy across various LLM models. The trends depicted in the graph further reinforce the significant improvements observed from baseline models, particularly with the adoption of CoT and structured prompting strategies.

The CoT prompts improve methodical thought and, therefore, improved accuracy, interpretation, and increased trust with transparency in the decision-making taken from the model predictions. It helps explain why post-hoc remains largely relevant for improving model performance while still supporting learning and decision quality in sensitive domains, among others, such as hate speech detection.

However, a few points will need some further investigation for optimization, for instance, poor quality of the dataset and computational limits. This will further be improved as even more sophisticated prompting approaches, including dynamic prompts from domain-specific data, and finer tuning with larger and more diverse datasets, will be used. The continuous learning approaches are also useful in the ongoing applicability and efficacy of the identification of hate speech.

TABLE I
MBZUAI-LaMini-T5-223M Model

Prompt	***Accuracy***	***Precision***	***Recall***	***F1***
default_0shot_no_cot	0.52	0.61	0.49	0.54
default_0shot_cot_expl_first	0.57	0.85	0.30	0.45
default_4shot_cot_ans_first	0.62	0.87	0.40	0.55
default_4shot_cot_expl_first	0.74	0.72	0.89	0.80

TABLE II
MBZUAI-LaMini-Neo-125M Model

Prompt	***Accuracy***	***Precision***	***Recall***	***F1***
default_0shot_no_cot	0.62	0.61	0.91	0.73
default_0shot_cot_expl_first	0.66	0.66	0.84	0.74
default_4shot_cot_ans_first	0.74	0.73	0.88	0.80
default_4shot_cot_expl_first	0.76	0.73	0.92	0.82

TABLE III
MBZUAI-LaMini-GPT-124M

Prompt	***Accuracy***	***Precision***	***Recall***	***F1***
default_0shot_no_cot	0.47	0.77	0.12	0.21
default_0shot_cot_expl_first	0.57	0.85	0.31	0.45
default_4shot_cot_ans_first	0.62	0.89	0.39	0.54
default_4shot_cot_expl_first	0.62	0.90	0.39	0.54

TABLE IV
MBZUAI-LaMini-Flan-T5-248M

Prompt	***Accuracy***	***Precision***	***Recall***	***F1***
default_0shot_no_cot	0.54	0.87	0.24	0.37
default_0shot_cot_expl_first	0.57	0.81	0.33	0.47
default_4shot_cot_ans_first	0.62	0.87	0.40	0.55
default_4shot_cot_expl_first	0.74	0.71	0.92	0.80

TABLE V
MBZUAI-LaMini-Cerebras-256M

Prompt	***Accuracy***	***Precision***	***Recall***	***F1***
default_0shot_no_cot	0.62	0.62	0.88	0.73
default_0shot_cot_expl_first	0.72	0.71	0.87	0.78
default_4shot_cot_ans_first	0.77	0.78	0.85	0.81
default_4shot_cot_expl_first	0.77	0.75	0.91	0.82

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:37:23 UTC from IEEE Xplore. Restrictions apply.

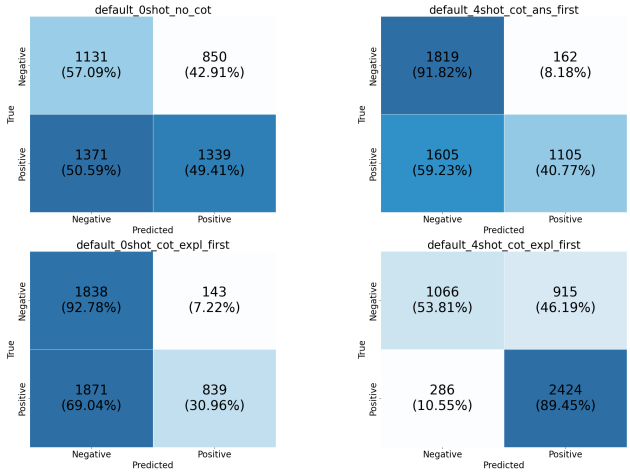


Fig. 6. Confusion Matrix LaMini-T5-223M

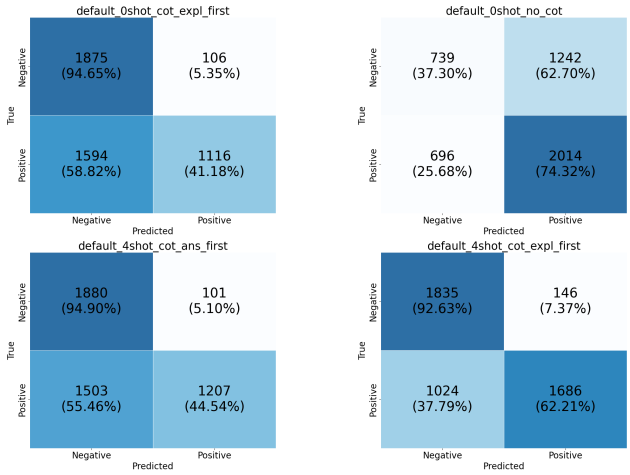


Fig. 7. Confusion Matrix for google-flan-t5-base

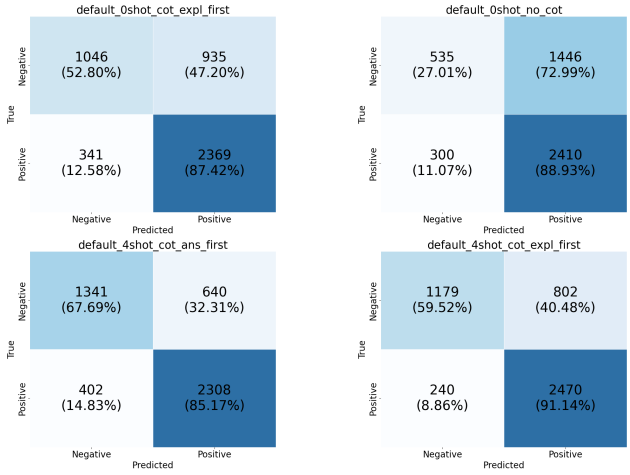


Fig. 8. Confusion Matrix for LaMini-Cerebras-256M

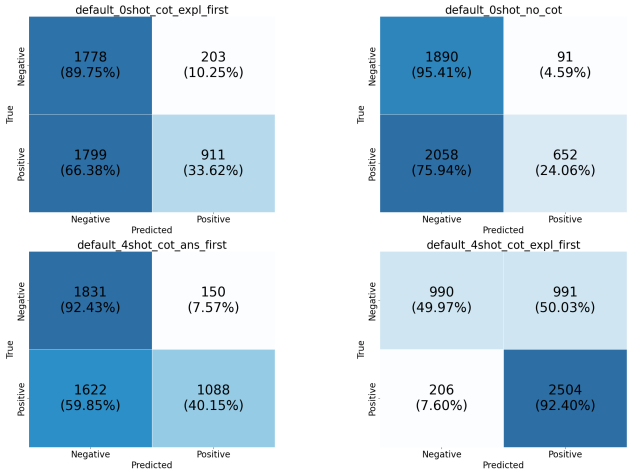


Fig. 9. Confusion Matrix for LaMini-Flan-T5-248M

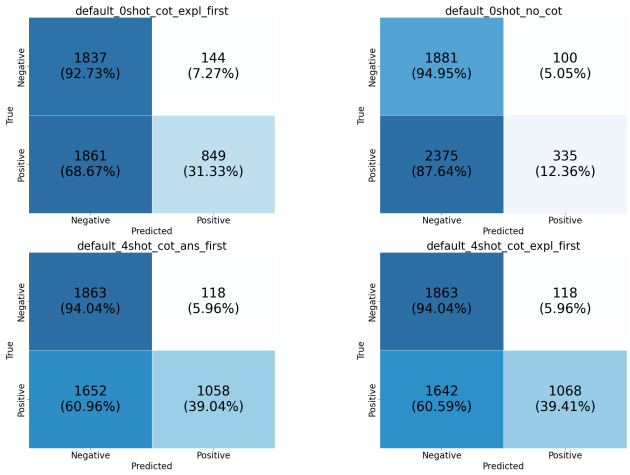


Fig. 10. Confusion Matrix for LaMini-GPT-124M

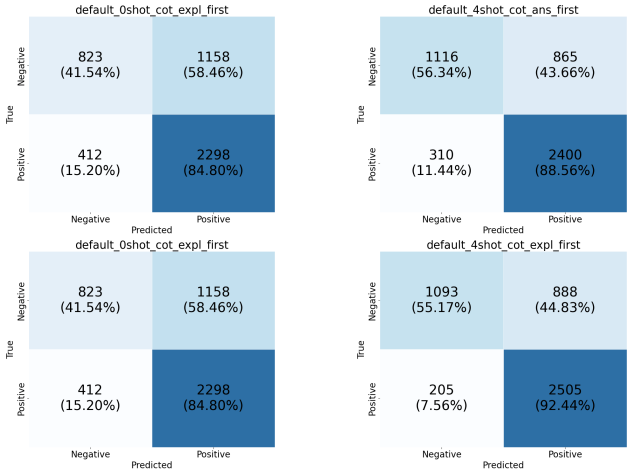


Fig. 11. Confusion Matrix for LaMini-Neo-125M

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:37:23 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
GOOGLE-FLAN-T5-BASE

Prompt	*Accuracy*	*Precision*	*Recall*	*F1*
default_0shot_no_cot	0.58	0.61	0.74	0.67
default_0shot_cot_expl_first	0.63	0.91	0.41	0.56
default_4shot_cot_ans_first	0.65	0.92	0.44	0.60
default_4shot_cot_expl_first	0.75	0.92	0.62	0.74

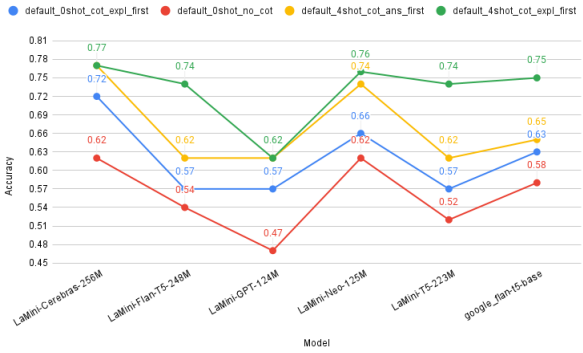


Fig. 12. Progression of accuracy across various prompt strategies for Lamini and Google-Flan-T5 Models.

V. Conclusion

This research study has analyzed the efficacy of Large Language Models (LLMs) in hate speech detection, exploring various prompting strategies and shot settings across LaMini and Google Flan-T5 models. The findings revealed insightful patterns in accuracy variations, highlighting the significant impact of contextual prompts and post hoc explanations on model performance. Analysing the patterns across different models and prompting strategies, we observe that contextual information, particularly when provided through CoT prompts, significantly enhances hate speech detection accuracy across different LLM models. Moreover, post hoc explanations emerged as a valuable tool for improving model interpretability and guiding the decision-making process, leading to substantial accuracy improvements in certain configurations.

Acknowledgment

This work was carried out as part of "Major Project" during eighth semester B Tech Computer Science and Engineering in partial fulfillment of the requirement for the award of Bachelor of Technology Degree in Computer Science and Engineering under Delhi Technological University.

References

[1] J. S. Vedeler, T. Olsen, and J. Eriksen, "Hate speech harms: a social justice discussion of disabled norwegians' experiences," *Disability amp; Society*, vol. 34, no. 3, p. 368–383, mar 2019. [Online]. Available: http://dx.doi.org/10.1080/09687599.2018.1515723

[2] S. T. Fiske, "Controlling other people: The impact of power on stereotyping." *American Psychologist*, vol. 48, no. 6, p. 621–628, 1993. [Online]. Available: http://dx.doi.org/10.1037/0003-066X.48.6.621

[3] M. Sap, S. Gabriel, L. Qin, and OTHERS, "Social bias frames: Reasoning about social and power implications of language," 2019. [Online]. Available: http://arxiv.org/pdf/1911.03891.pdf

[4] F. Huang, H. Kwak, and J. An, "Is chatgpt better than human annotators? potential and limitations of chatgpt in explaining implicit hate speech," in *Companion Proceedings of the ACM Web Conference 2023*, ser. WWW '23. ACM, apr 2023. [Online]. Available: http://dx.doi.org/10.1145/3543873.3587368

[5] A. K. Lampinen, I. Dasgupta, S. C. Y. Chan, and OTHERS, "Can language models learn from explanations in context," 2022.

[6] S. Krishna, J. Ma, D. Slack, and OTHERS, "Post hoc explanations of language models can improve language models," 2023. [Online]. Available: http://arxiv.org/pdf/2305.11426.pdf

[7] M. Suzgun, N. Scales, N. Schärli, and OTHERS, "Challenging big-bench tasks and whether chain-of-thought can solve them," 2022. [Online]. Available: http://arxiv.org/pdf/2210.09261.pdf

[8] T. Kojima, S. S. Gu, M. Reid, and OTHERS, "Large language models are zero-shot reasoners," 2022. [Online]. Available: http://arxiv.org/pdf/2205.11916.pdf

[9] A.-M. Founta, C. Djouvas, D. Chatzakou, and OTHERS, "Large scale crowdsourcing and characterization of twitter abusive behavior," 2018. [Online]. Available: http://arxiv.org/pdf/1802.00393.pdf

[10] R. Martins, M. Gomes, J. J. Almeida, P. Novais, and P. Henriques, "Hate speech classification in social media using emotional analysis," in *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*. IEEE, oct 2018. [Online]. Available: http://dx.doi.org/10.1109/bracis.2018.00019

[11] B. Mathew, P. Saha, S. M. Yimam, C. Biemann, P. Goyal, and A. Mukherjee, "Hatexplain: A benchmark dataset for explainable hate speech detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 17, p. 14867–14875, may 2021. [Online]. Available: http://dx.doi.org/10.1609/aaai.v35i17.17745

[12] J. Devlin, M.-W. Chang, K. Lee, and OTHERS, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018. [Online]. Available: http://arxiv.org/pdf/1810.04805.pdf

[13] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.

[14] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners," 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:160025533

[15] M. ElSherief, C. Ziems, D. Muchlinski, V. Anupindi, J. Seybolt, M. De Choudhury, and D. Yang, "Latent hatred: A benchmark for understanding implicit hate speech," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2021. [Online]. Available: http://dx.doi.org/10.18653/v1/2021.emnlp-main.29

[16] L. Li, L. Fan, S. Atreja, and OTHERS, ""hot" chatgpt: The promise of chatgpt in detecting and discriminating hateful, offensive, and toxic comments on social media," 2023. [Online]. Available: http://arxiv.org/pdf/2304.10619.pdf

[17] W. Ling, D. Yogatama, C. Dyer, and OTHERS, "Program induction by rationale generation : Learning to solve and explain algebraic word problems," 2017. [Online]. Available: http://arxiv.org/pdf/1705.04146.pdf

[18] K. Cobbe, V. Kosaraju, M. Bavarian, and OTHERS, "Training verifiers to solve math word problems," 2021. [Online]. Available: http://arxiv.org/pdf/2110.14168.pdf

[19] M. Nye, A. J. Andreassen, G. Gur-Ari, and OTHERS, "Show your work: Scratchpads for intermediate computation with language models," 2021. [Online]. Available: http://arxiv.org/pdf/2112.00114.pdf

[20] T. B. Brown, B. Mann, N. Ryder, and OTHERS, "Language models are few-shot learners," 2020. [Online]. Available: http://arxiv.org/pdf/2005.14165.pdf

[21] A. Nasir, A. Sharma, and K. Jaidka, "Llms and finetuning: Benchmarking cross-domain performance for hate speech detection," 2023. [Online]. Available: http://arxiv.org/pdf/2310.18964

[22] F. M. Plaza-del arco, D. Nozza, and D. Hovy, "Respectful or toxic? using zero-shot learning with language models to detect hate speech," in *The 7th Workshop on Online Abuse and Harms (WOAH)*. Association for Computational Linguistics, 2023. [Online]. Available: http://dx.doi.org/10.18653/v1/2023.woah-1.6

[23] J. Wei, X. Wang, D. Schuurmans, and OTHERS, "Chain-of-thought prompting elicits reasoning in large language models," 2022. [Online]. Available: http://arxiv.org/pdf/2201.11903.pdf

[24] M. Chen, J. Tworek, H. Jun, and OTHERS, "Evaluating large language models trained on code," 2021. [Online]. Available: http://arxiv.org/pdf/2107.03374.pdf

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:37:23 UTC from IEEE Xplore. Restrictions apply.

[25] M. Wu, A. Waheed, C. Zhang, and OTHERS, "Lamini-lm: A diverse herd of distilled models from large-scale instructions," 2023. [Online]. Available: http://arxiv.org/pdf/2304.14402.pdf
[26] Y. Wang, Y. Kordi, S. Mishra, and OTHERS, "Self-instruct: Aligning language models with self-generated instructions," 2022. [Online]. Available: http://arxiv.org/pdf/2212.10560.pdf
[27] V. Sanh, A. Webson, C. Raffel, and OTHERS, "Multitask prompted training enables zero-shot task generalization," 2021. [Online]. Available: http://arxiv.org/pdf/2110.08207.pdf
[28] S. Longpre, L. Hou, T. Vu, and OTHERS, "The flan collection: Designing data and methods for effective instruction tuning," 2023. [Online]. Available: http://arxiv.org/pdf/2301.13688.pdf
[29] C. Raffel, N. Shazeer, A. Roberts, and OTHERS, "Exploring the limits of transfer learning with a unified text-to-text transformer," 2019. [Online]. Available: http://arxiv.org/pdf/1910.10683.pdf
[30] H. W. Chung, L. Hou, S. Longpre, and OTHERS, "Scaling instruction-finetuned language models," 2022. [Online]. Available: http://arxiv.org/pdf/2210.11416.pdf
[31] N. Dey, G. Gosal, Zhiming, and OTHERS, "Cerebras-gpt: Open compute-optimal language models trained on the cerebras wafer-scale cluster," 2023. [Online]. Available: http://arxiv.org/pdf/2304.03208.pdf
[32] L. Gao, S. Biderman, S. Black, and OTHERS, "The pile: An 800gb dataset of diverse text for language modeling," 2020. [Online]. Available: http://arxiv.org/pdf/2101.00027.pdf
[33] H. Gonen, S. Iyer, T. Blevins, and OTHERS, "Demystifying prompts in language models via perplexity estimation," 2022. [Online]. Available: http://arxiv.org/pdf/2212.04037.pdf

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:37:23 UTC from IEEE Xplore. Restrictions apply.

Uncorrected Proof

Water Supply

© 2024 The Authors

Water Supply Vol 00 No 0, 1 doi: 10.2166/ws.2024.172

Impact of flexible vegetation on hydraulic characteristics in compound channels with converging floodplains

Noopur Awasthi *, Munendra Kumar and Ritu Raj

Department of Civil Engineering, Delhi Technological University, Delhi 110042, India
*Corresponding author. E-mail: noopurawasthi_2k21phdce502@dtu.ac.in

NA, 0000-0001-9033-2567

ABSTRACT

Floodplains in natural rivers are curved and narrow, rather than following a straight path and the presence of vegetation and varying depth of flow affects the flow parameters of the channels. The purpose of this study is to investigate the effects of the convergence of the channel and relative depth of flow on the depth-averaged velocity (DAV) and boundary shear stress (BSS) of a converging compound channel. The effect of vegetation on DAV and BSS was also deliberated. To examine the effect of vegetation, synthetic grass 8 cm in height was used as a flexible vegetation to depict the natural conditions of the rivers. The DAV was calculated using an acoustic Doppler velocimeter (ADV) and the measurement of BSS was done using the Preston tube technique. After analysing the data, it was revealed that as the convergence of floodplains increases, maximum DAV increases. The presence of vegetation reduces the velocity specifically at lower flow depths. BSS increases with the convergence as well as vegetation. This research provides valuable data and results which will be helpful in better flood management, river restoration, and designing effective hydraulic structures.

Key words: boundary shear stress, compound channel, convergence, depth-averaged velocity, flexible vegetation, relative depth

HIGHLIGHTS

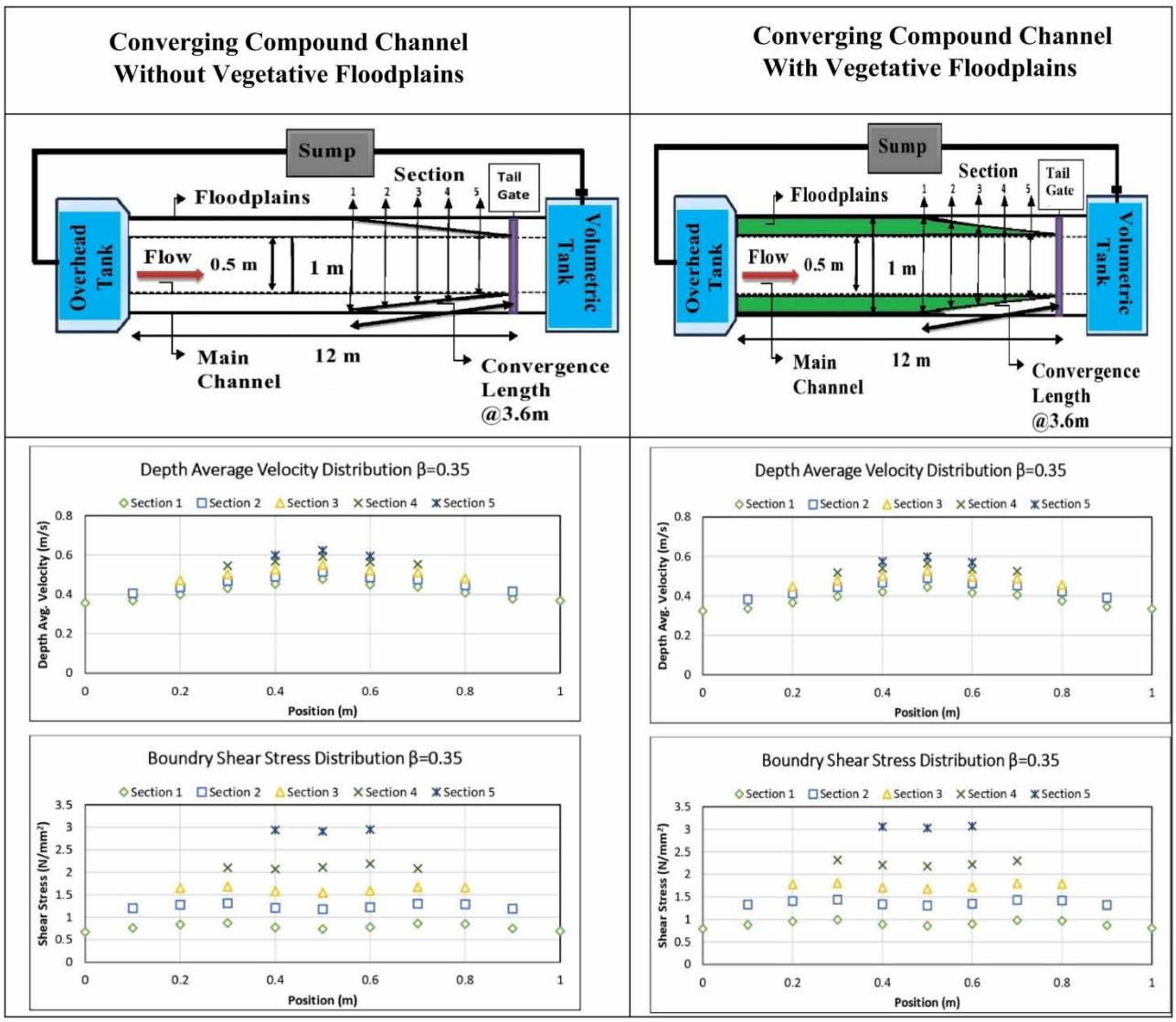
- The study examines the influence of channel convergence across various flow depths on depth-averaged velocity and boundary shear stress within a compound channel.
- Synthetic grass (flexible vegetation) is utilized to mimic natural vegetation, offering insights into vegetation's influence on flow parameters.
- The study's findings provide valuable insights for enhancing flood management, river restoration, and hydraulic structure design.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

GRAPHICAL ABSTRACT



NOTATIONS

The following symbols are used in this article:

B width of compound channel
b width of main channel
h height of main channel
H depth of flow
β relative flow depth $[(H - h)/H]$
θ converging angle
U_d depth-averaged velocity
Q discharge of compound channel

INTRODUCTION

The acceleration of global warming brought on by human activities has hastened climate change, which has impacted precipitation patterns and made extreme weather events more frequent. These changes are increasing the likelihood of floods in many locations, putting already vulnerable people at risk. Among all the natural disasters that a country faces, river

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

floods are the most frequent and often devastating (Alam & Muzzammil 2011). With the substantial changes in the global climate, growing urbanization, and land use, the dynamics of flooding are changing in ways that are complicated and unprecedented. Understanding river hydraulics is a crucial initial phase in the examination and management of floods. Determining flow discharge in rivers is a crucial aspect of designing river engineering projects (Azamathulla & Zahiri 2012). When assessing flood damages, it is essential to analyse the flow parameters in both the main channel and the floodplains (Pandey *et al.* 2022). By delving into these flow parameters, researchers and flood management authorities can gain a deeper understanding of the mechanisms at play, facilitating a more accurate assessment of the potential damages and aiding in the development of effective mitigation strategies.

Due to increased urbanization and changes in land use patterns, many natural watercourses differ from the uniform prismatic shape, although prismatic compound channels with uniform cross-sections have been the subject of several classic hydraulic research (Khatua *et al.* 2012; Proust *et al.* 2013). To handle these systems effectively, one must have a clearer understanding of their inherent non-prismatic complexity. The previous studies (Bousmar *et al.* 2004; Proust *et al.* 2006; Rezaei & Knight 2011; Yonesi *et al.* 2013; Das & Khatua 2018) reveal that flow parameters in compound channels with non-prismatic floodplains are considerably more complicated than that in prismatic channels. The investigation of flow in prismatic channels is a well-explored topic in current research. In the design phase, crucial parameters such as turbulent characteristics, conveyance capacity, mean flow pattern, distribution of depth-averaged velocity (DAV), secondary flow features, and boundary shear stress (BSS) play a significant role. Hydraulic engineers often rely on experimental research to gain insights into the complex flow parameters involved when examining turbulent flows in open channels. This complexity escalates when the floodplain is non-prismatic inducing mass transfer. The flow velocity decreases on spreading floodplains, while it increases on converging floodplains due to the convergence of channel shape, as observed in studies conducted by Rezaei (2006). Rezaei & Knight (2011) conducted experiments on compound channels with symmetrically narrowing floodplains, emphasizing the geometric transfer of momentum and the resulting additional head losses.

Many natural rivers display a two-stage geometry, and floodplains feature diverse roughness elements, including various sizes of gravel, vegetation, and combinations of these elements. Analysing the flow parameters becomes more challenging as the flow enters its floodplains.

Vegetation, including submerged, emergent, and floating types, significantly influences the flow characteristics within a channel. The introduction of vegetation elements adds complexity to the flow structure in non-prismatic compound channels. Brito *et al.* (2016) and Yang *et al.* (2007) conducted experimental studies to explore how vegetation impacts the three-dimensional flow properties. The presence of vegetation in floodplains not only alters the hydraulic capacity of compound channels but also impacts river ecosystems, influencing flow attributes like velocity distribution, turbulence patterns, vortices, and the exchange of mass and momentum between vegetated and non-vegetated zones (Nezu & Sanjou 2008; Uotani *et al.* 2014). Numerical simulations have been employed by numerous scholars, including Zhang *et al.* (2017) and Koftis & Prinos (2018) to investigate the characteristics of turbulent flow patterns within vegetated channels. Most of the research on vegetated floodplains has primarily concentrated on straight compound channels, while there is limited literature on compound channels with vegetation on diverging and converging floodplains (Mehrabani *et al.* 2020; Rahim *et al.* 2023). In previous studies, rigid vegetation models have been used by several researchers in their studies of compound channels (Rahim *et al.* 2022; Rao *et al.* 2022; Zhang & Hu 2023). However, it is important to note that in the natural river systems, flexible vegetation is prevalent. The fundamental distinction between rigid and flexible vegetation types can have significant implications for our understanding of how compound channels interact with their surrounding environments. It is crucial to investigate the impact and dynamics of flexible vegetation within compound channels. This article aims to bridge the gap between theoretical research and the actual conditions present in natural river ecosystems. Moreover, flexible artificial plants have been used to simulate floodplain vegetation in compound channels, and flume experiments have produced results similar to field observations (Huai *et al.* 2013; Liu *et al.* 2016). Previous studies have identified several unexplored aspects and gaps regarding vegetated floodplains in non-prismatic compound channels.

This article investigates the flow parameters in a converging compound channel with vegetative floodplains, the focus is on comprehending the flow physics under converging floodplains, considering flexible vegetation at various flow depths.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

EXPERIMENTAL SET-UP

Experiments were conducted at the Hydraulics Laboratory of the Department of Civil Engineering, Delhi Technological University, using a symmetrical compound channel of dimensions 12 m long, 1 m wide, and 0.8 m deep. The cross-sectional width of the main channel (b) was 0.5 m and the depth (h) was 0.25 m. The converging segment of the channel had a convergence angle (θ) of 4°. The whole set-up was constructed with brick masonry and plastered with concrete to provide a smooth surface finish. The compound channel comprised a 6 m long prismatic section and a 3.6 m long non-prismatic section and a downstream portion as presented in Figure 1. To represent the flexible vegetation in the floodplains, the flume's bottom was covered with synthetic grass, with an average height of 8 cm as presented in Figure 2.

METHODOLOGY

The flume was operated at six different relative flow depths ($\beta = 0.20, 0.30, 0.35, 0.40, 0.45, 0.50$). The discharge in the channel varied with the relative depth.

$$\text{Relative depth } (\beta) = \frac{H - h}{H} \tag{1}$$

where H represents the water height at a specific section and h indicates the height of water in the main channel.

The relation between discharge (Q) and relative depth (β) is presented in Figure 3. Totally five sections termed Section 1, Section 2, Section 3, Section 4, and Section 5 were selected for the study as represented in Figure 1. Section 1 and Section 5 denote the start and end of channel convergence, while Section 2, Section 3, and Section 4, positioned equidistantly, serve as intermediate sections. The longitudinal distance between every section is constant. Velocity and shear stress were measured

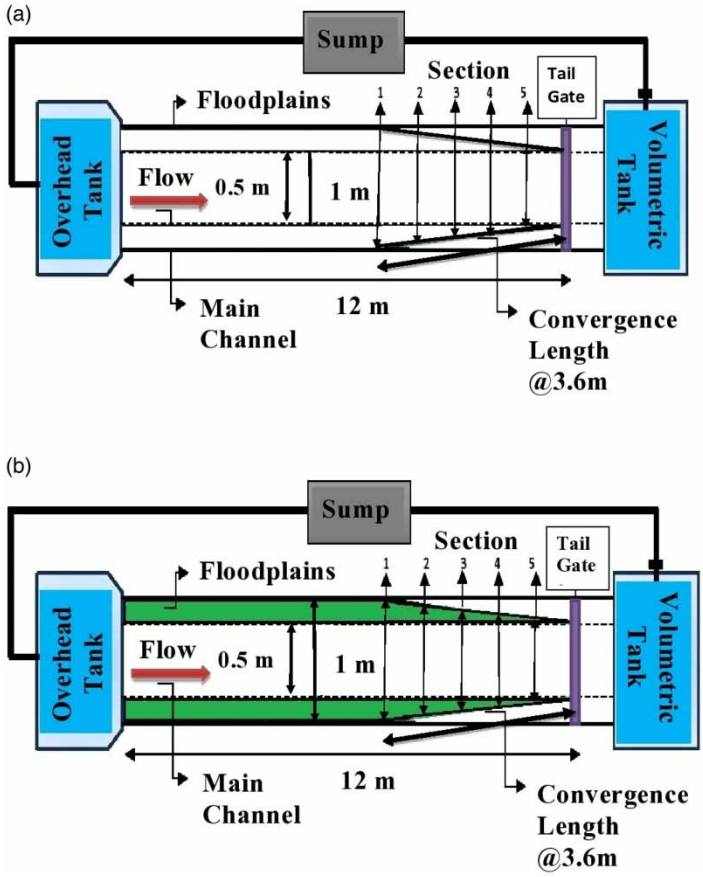


Figure 1 | (a) Plan view of experimental set-up without vegetation. (b) Plan view of experimental set-up with vegetation.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof



Figure 2 | (a) Laboratory set-up of converging channel without vegetation. (b) Laboratory set-up of converging channel with vegetation.

at every non-prismatic section for each relative depth. Detailed information about measuring of velocity and shear stress is given in further sections.

FLOW CONDITIONS AND MEASUREMENTS

The water was drawn from an underneath sump and pumped to an overhead tank in the testing channel. A volumetric tank fitted with a v-notch was exercised for measuring the discharge from the experimental compound channel. The v-notch was calibrated for this specific purpose, and the water was subsequently returned to the sump located underneath. To control the water surface profile and enforce a particular depth of flow in the flume portion, a tailgate was installed at the downstream end of the flume.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

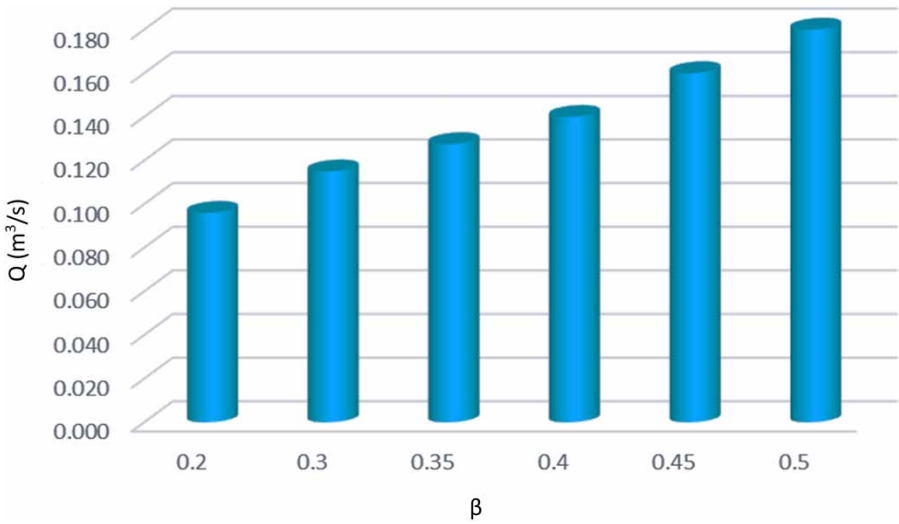


Figure 3 | Discharge vs. relative depth.

A point gauge with a precision of 0.10 mm was utilized to measure the water surface profile at distances of 1.0 and 0.30 m in the prismatic length and non-prismatic length of the compound channel, respectively. The point velocity at the pre-defined points along the wetted perimeter was determined using an acoustic Doppler velocimeter (ADV) at vertical intervals of 2.5 cm and horizontal intervals of 10 cm. The data obtained from the ADV underwent filtering using the ADV Horizon software.

The lateral variations in BSS were additionally gauged using a Preston tube featuring a 4.77 mm outer diameter at the identical sections where velocity distributions were assessed as in Figure 4. For measuring the pressure difference, a digital manometer was used, and Patel's (1965) calibration equations were used for calculating shear stress values.

Depth-averaged velocity

In this study, velocity measurements were performed using both Pitot tubes and a 16-MHz Micro-ADV. The ADV is unable to calculate velocity within 50 mm of the free surface and bed due to limitations of the instrument. To address this limitation, a standard Pitot tube with a 5 mm outer diameter was used in conjunction with a digital manometer to measure point velocity readings at specified locations within the upper and bottom 50 mm regions of the channel. Point velocities were measured at various points within pre-defined grids across the channel cross-sections to ensure comprehensive coverage of the entire flow area.

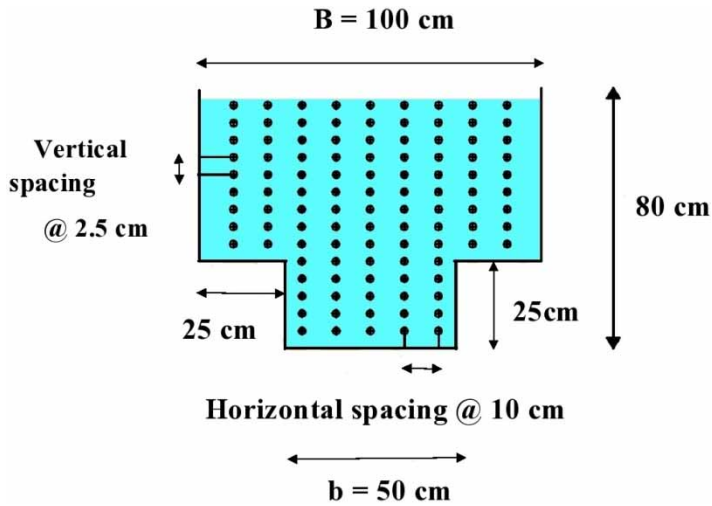


Figure 4 | Grid points on cross-section of compound channel.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

For every experimental configuration, velocity calculations were conducted at six different discharges, corresponding to six relative depths ($\beta = 0.20, 0.30, 0.35, 0.4, 0.45, 0.50$), at five different sections. The depth-averaged velocity (U_d) was computed using the following equation (Naghavi *et al.* 2023):

$$U_d = \frac{1}{H}\int_0^H u(z)dz \tag{2}$$

where U_d is the depth-averaged velocity; $u(z)$ denotes local velocity at z depth; and H is the flow depth.

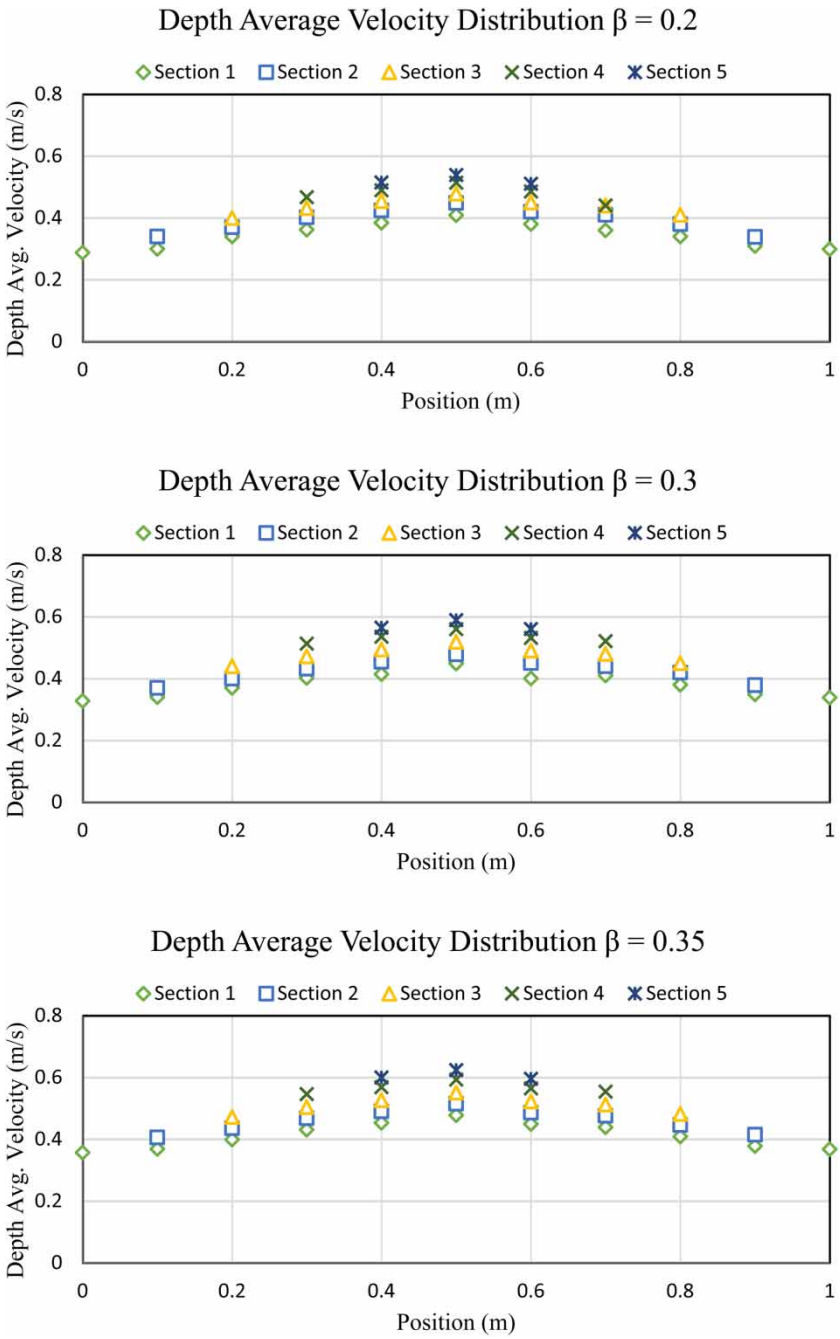


Figure 5 | DAV distribution at all six relative depths ($\beta = 0.2, 0.3, 0.35, 0.4, 0.45, 0.5$) without vegetation. (*continued.*).

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

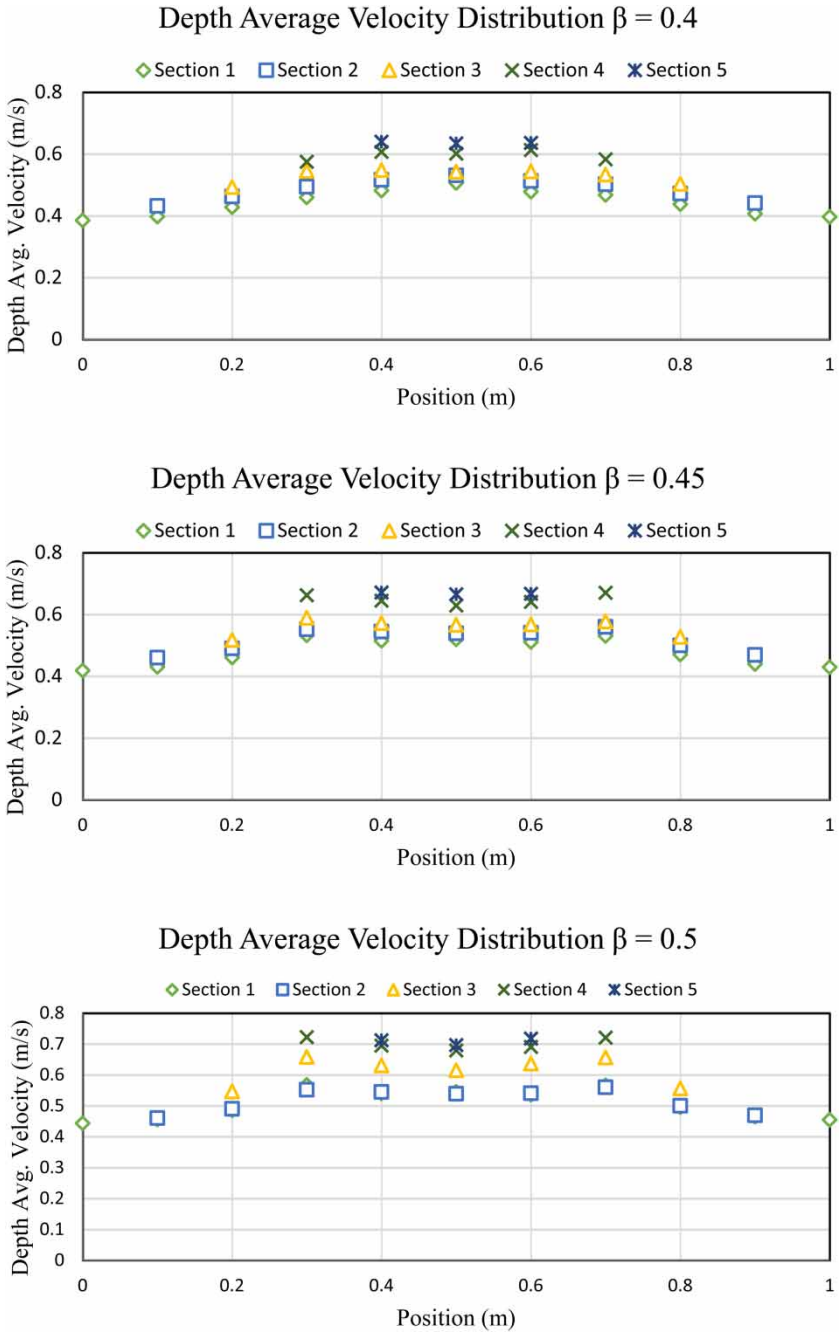


Figure 5 | Continued.

The parameter U_d is of crucial importance in compound channel flow studies. Accurate measurement of U_d and its distribution across the flow section is essential for understanding the behaviour of compound channels under varying relative depths.

Boundary shear stress

Explorations into shear stress within open channel flow carry extensive implications, impacting sediment transport, channel migration, calculation of energy losses, and momentum transfer. The Preston tube technique stands out as a widely employed indirect approach for gauging shear stress in measurements. In this study, local BSS measurements were conducted using Preston's technique (1954) in conjunction with Patel's (1965) calibration curves. The relation between BSS and differential

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

pressure was defined by Preston as follows:

$$\frac{\Delta p d^2}{\rho \nu^2} = F\left(\frac{d^2 \tau_w}{\rho \nu^2}\right) \tag{3}$$

where Δp is the differential pressure; F is an empirical function; d is the Preston tube outer diameter; ρ is the fluid density; ν is the fluid kinematic viscosity; and τ_w is the boundary shear stress.

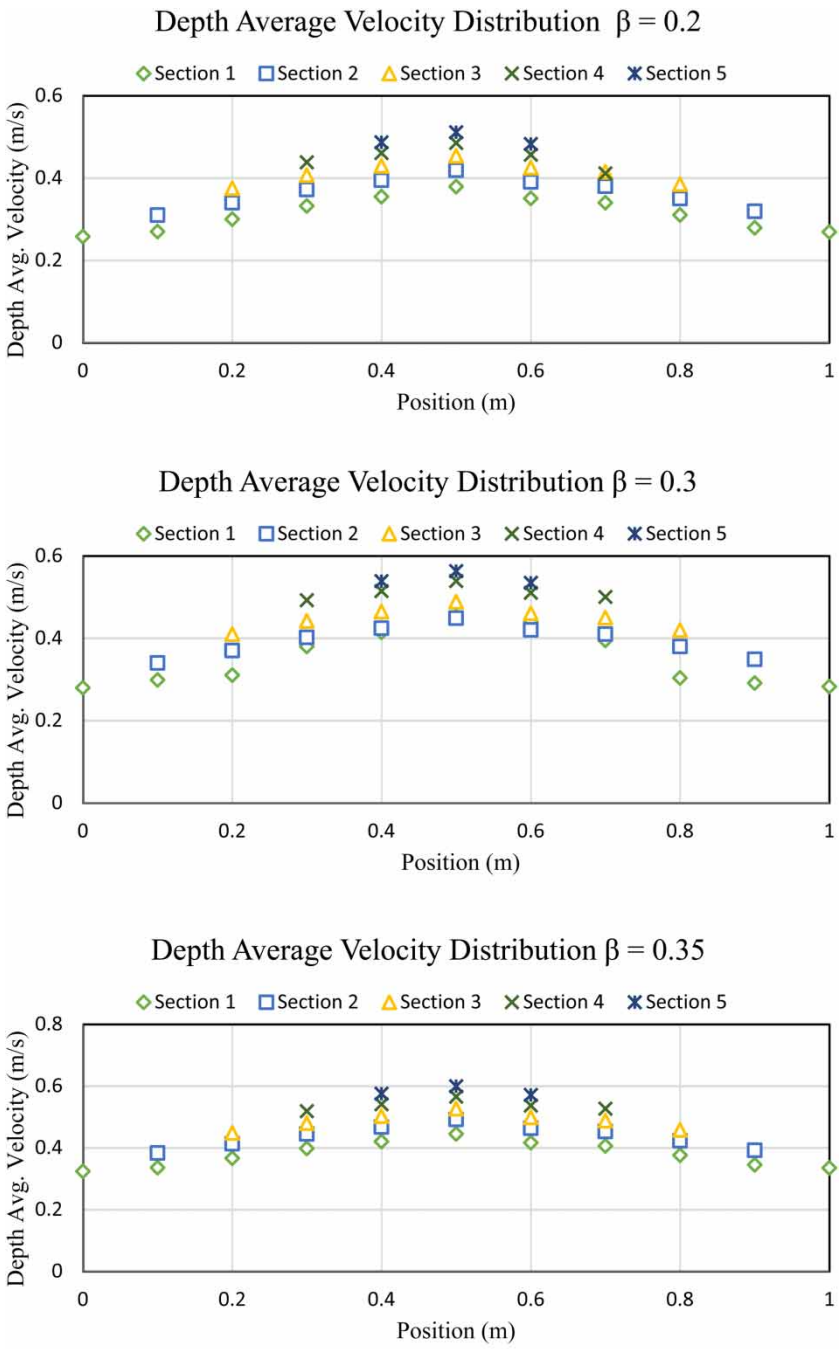


Figure 6 | DAV distribution at all six relative depths ($\beta = 0.20$, 0.30, 0.35, 0.40, 0.45, 0.50) with vegetation. (*continued.*).

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

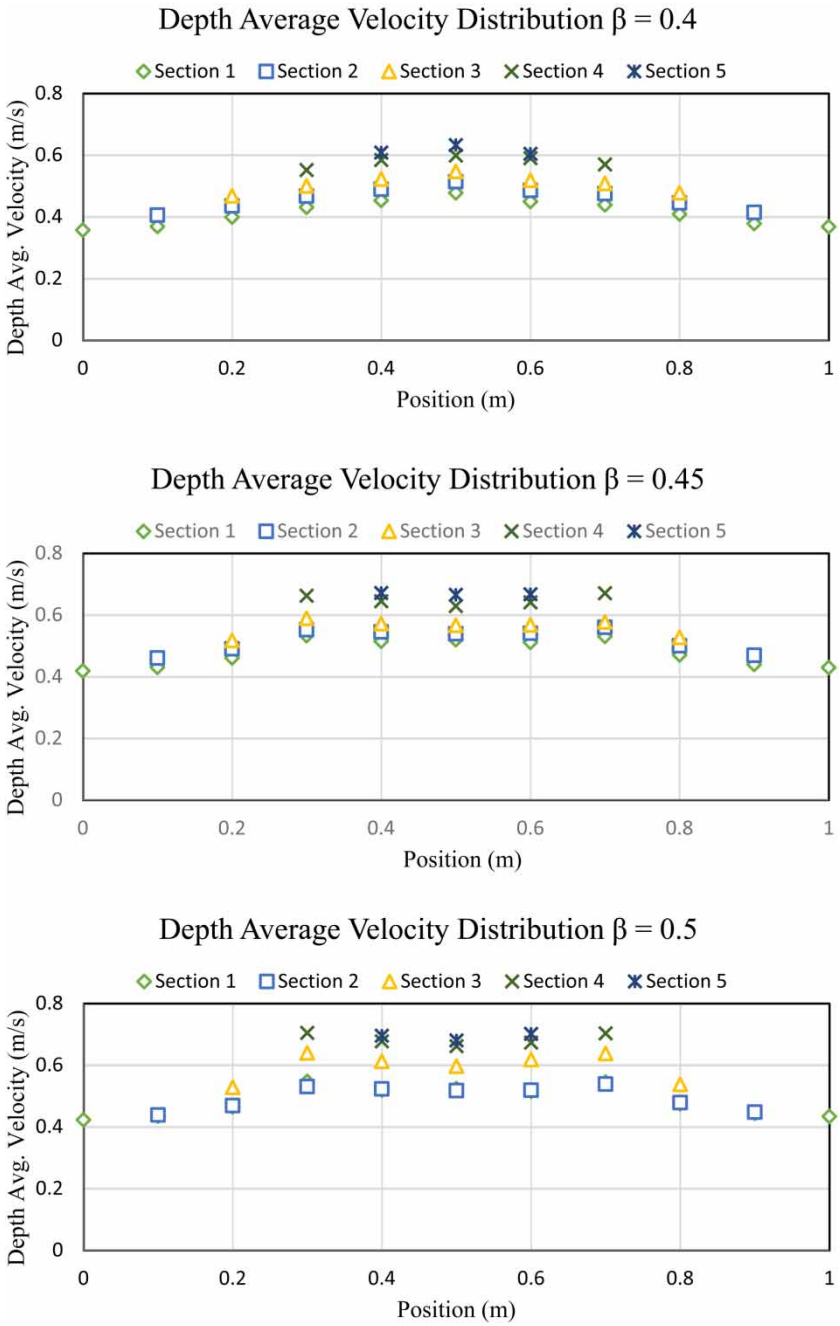


Figure 6 | Continued.

Patel (1965) developed precise calibration curves for the Preston tube based on two non-dimensional parameters, y^* and x^*, facilitating the conversion of pressure readings into BSS. The calibration equations are as follows:

$$x^* = \log_{10}\left(\frac{\Delta p d^2}{4\rho\nu^2}\right) \tag{4}$$

$$y^* = 0.5x^* + 0.037 \quad \text{for } 0 < y^* < 1.5 \text{ and } 0 < x^* < 2.9 \tag{5}$$

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

$$y^* = 0.8287 - 0.1381x^* + 0.1437x^{*2} - 0.0060\,x^{*3} \quad \text{for } 1.5 < y^* < 3.5 \tag{6}$$

$$x^* = y^* + 2\log_{10}(1.95y^* + 4.10) \quad \text{for } 3.5 < y^* < 5.3 \tag{7}$$

$$y^* = \left(\frac{\pi d^2}{4\rho\nu^2}\right) \tag{8}$$

where x^* is logarithm of the dimensionless pressure difference and y^* is logarithm of the dimensionless shear stress

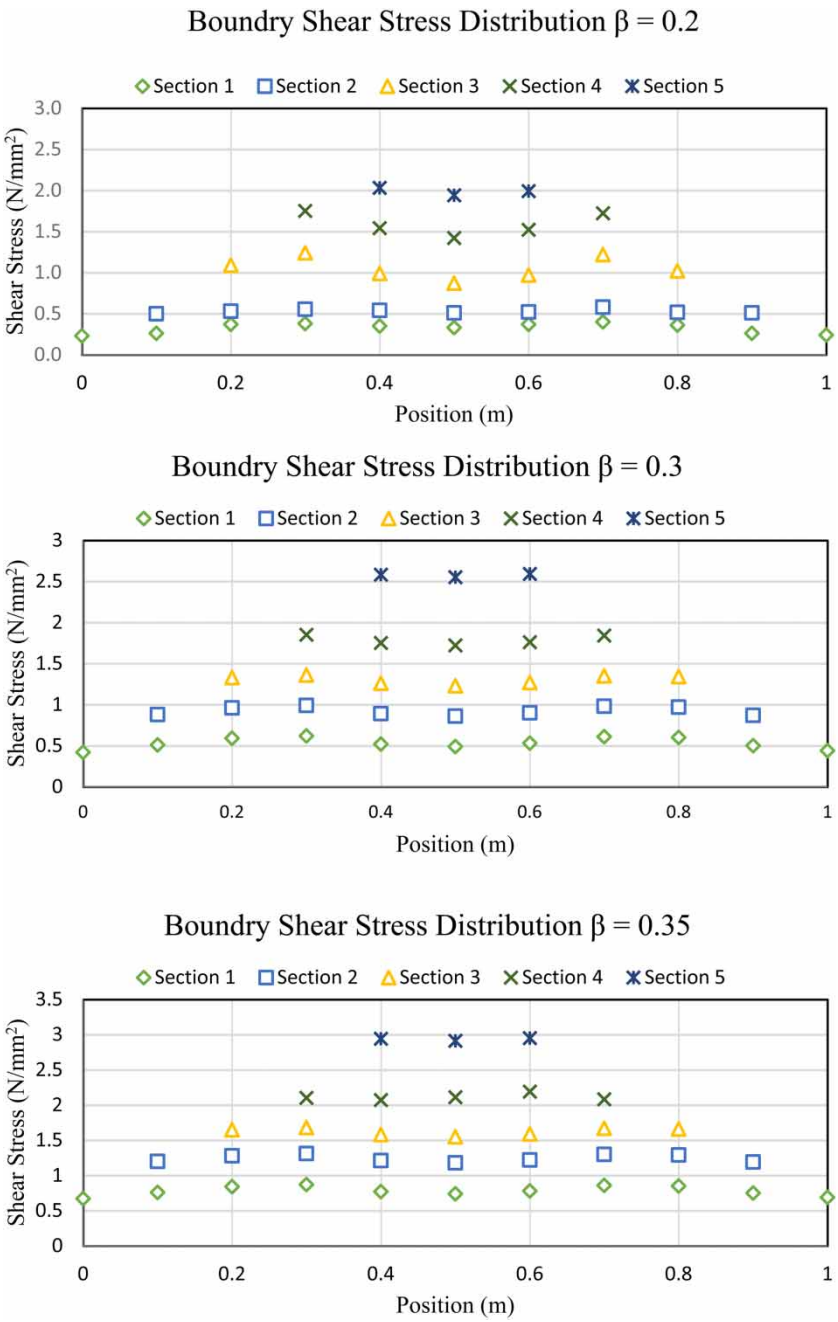


Figure 7 | BSS distribution at all six relative depths ($\beta = 0.2$, 0.3, 0.35, 0.4, 0.45, 0.5) without vegetation. (*continued.*).

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

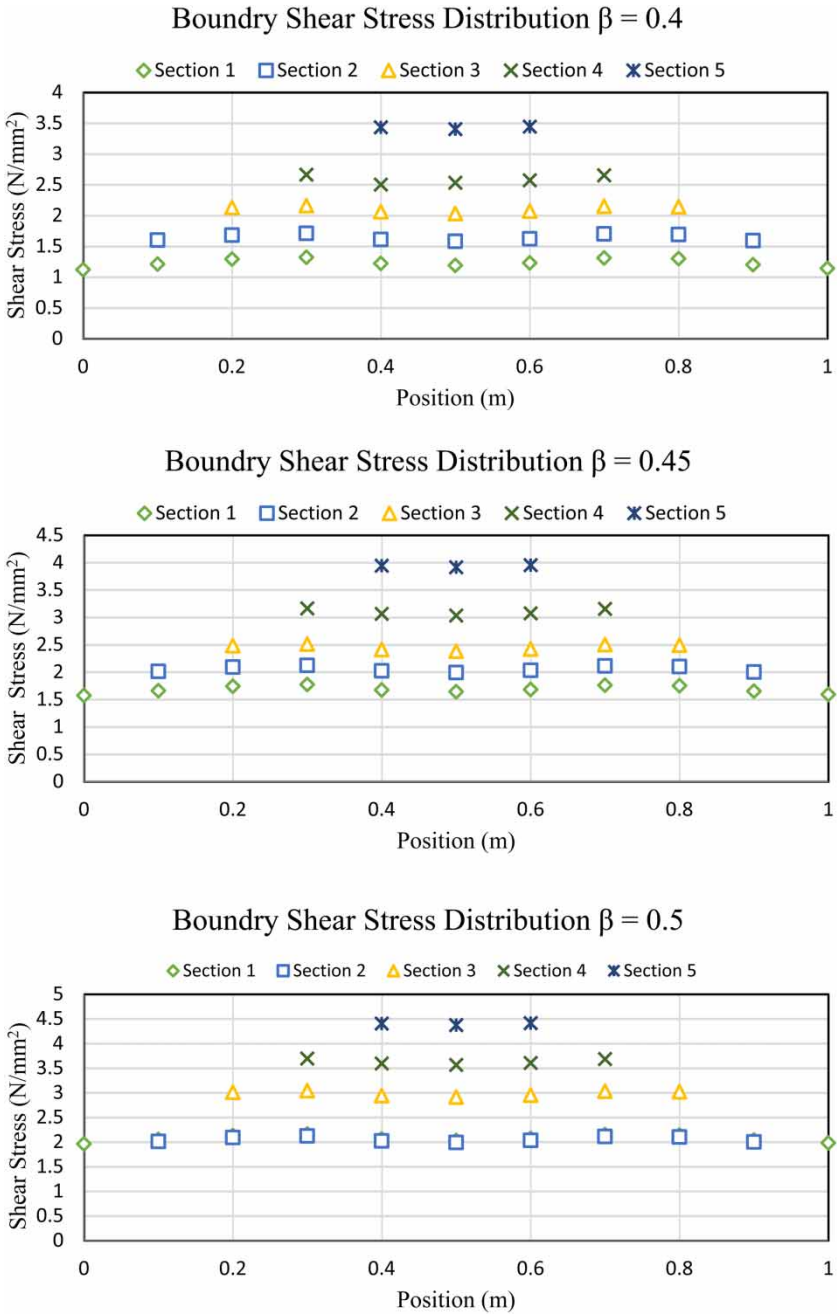


Figure 7 | Continued.

Shear stress assessments were carried out in each of the five sections of the channel. Pressure readings were obtained using a Preston tube positioned at previously defined grid points within the channel, oriented towards the flow. The differential pressure was measured using a digital manometer connected to the Preston tube.

RESULTS AND ANALYSIS

The flow is found to be non-uniform in the converging length of the channel. Subcritical flow conditions were attained under various conditions of the compound channel with a longitudinal bed slope of 0.001. The results of the DAV and BSS are discussed further in this section.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Depth-averaged velocity

The distribution of velocity across all cross-sections is illustrated in Figure 5 for all six relative depths ($\beta = 0.20$, 0.30, 0.35, 0.40, 0.45, 0.50). The observed rise in maximum average depth velocity across the channel coincides with the convergence of floodplains. There is a considerable increase in the velocity of the second half of the converging reach. The flow becomes increasingly restricted as a result of converging floodplains thus reducing the channel width, which results in momentum

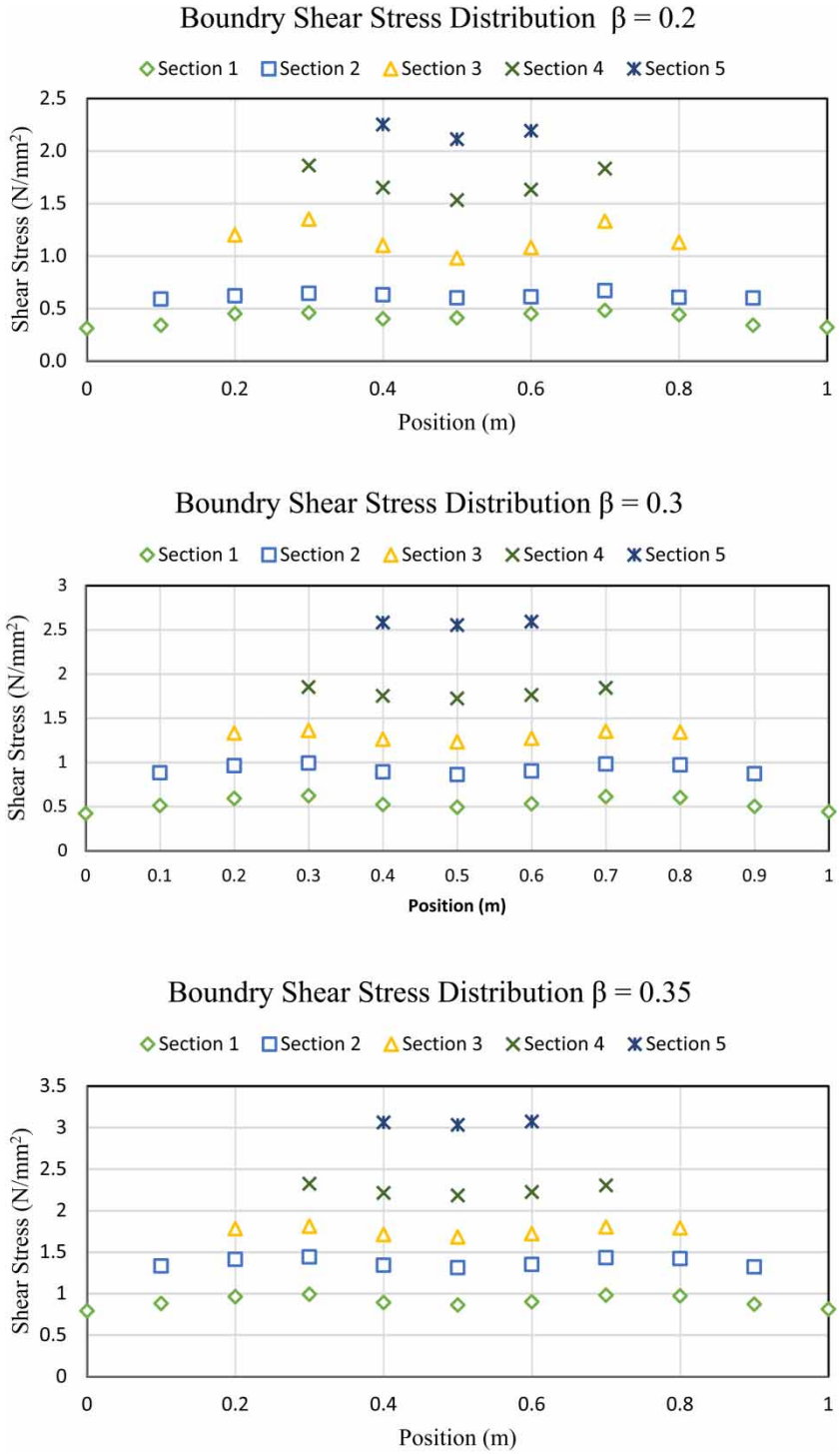


Figure 8 | BSS distribution at all six relative depths ($\beta = 0.2$, 0.3, 0.35, 0.4, 0.45, 0.5) with vegetation. (*continued.*).

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

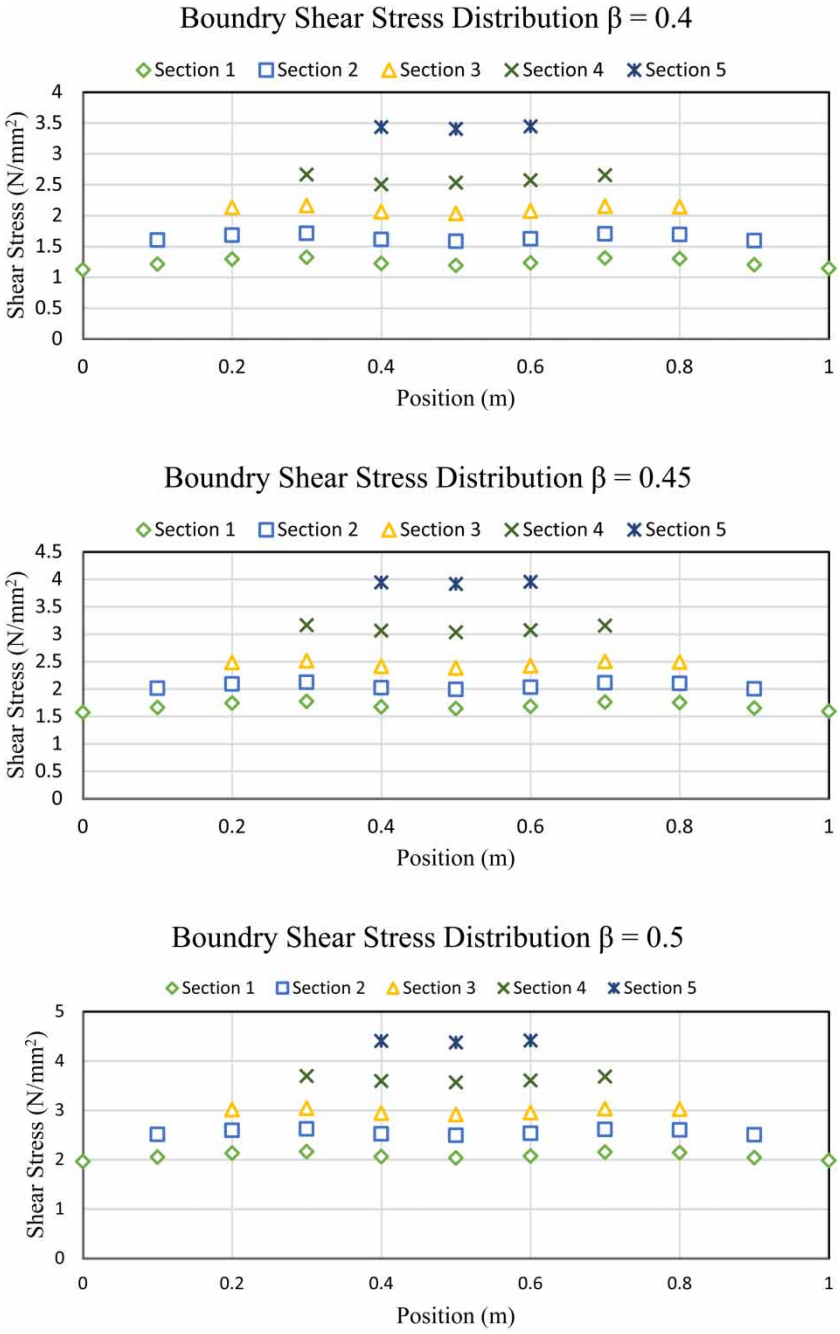


Figure 8 | Continued.

transfer. The increased velocity indicates the concentration of flow, resulting in a more efficient conveyance of water through the channel. Foreseeing the effect of channel geometry on overall velocity patterns and hydraulic efficiency in non-prismatic channels with converging floodplains becomes essential when understanding this phenomenon. Naik *et al.* (2018) conducted an experiment of overbank flow in a converging channel, exploring three convergent angles ($\theta = 5°$, $\theta = 9°$, and $\theta = 12.38°$). The study revealed that one of the results of the convergence on flow behaviour is an increase in velocity along the converging segment of the flume, aligning with the findings of the current research.

The observations of the study depicted in Figure 5 highlight that the highest velocity is typically concentrated in the central region of the main channel for the lower relative depths. The central concentration of maximum velocity in the main channel

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

can be related to the fact that the highest flow velocities arise from diminished frictional resistance along this central axis. For higher relative depths, the maximum velocity shifts towards the floodplains.

Figure 6 illustrates how the flow velocity across the channel is significantly reduced when there is the presence of flexible vegetation irrespective of the relative depth in all cross-sections. It can be observed from Figure 6 that the maximum DAV in the main channel is in the centre for lower flow depths, whereas, when the flow depth increases, the maximum centre velocity shifts towards floodplains. Maximum DAV in the channel with a vegetated floodplain is lower compared with a scenario without vegetation due to the frictional resistance by the vegetation. Moreover, at higher flow depths, the decrease in flow velocity in the main channel and floodplains due to vegetation is less pronounced compared with lower flow depths. This effect of the floodplain vegetation at higher relative depths was never shown in previous studies.

Boundary shear stress

Figure 7 exhibits the experimental findings of the BSS distributions in all cross-sections without vegetation for the converging flood plain of angle 4° for all six relative depths ($\beta = 0.20$, 0.30, 0.35, 0.40, 0.45, and 0.50). The BSS distribution in all sections is logically symmetrical in these figures and progressively increases from $\beta = 0.2$ to $\beta = 0.5$. The minimum BSS is at the floodplains, whereas the maximum BSS is at the edge of the floodplain and main channel which is relatively higher than centreline shear stress along the channel as the floodplains converge. Notably, in the second half of the converging reach, there is a significant surge in BSS. This surge is attributed to the mass transfer from the floodplains into the main channel at the end of convergence.

Experimental results of the BSS distribution in all sections with vegetation for the converging flood plain of angle 4° for all relative depths 0.20, 0.30, 0.35, 0.40, 0.45, and 0.50, are presented in Figure 8. It shows similar distributions as without vegetation, but the flexible vegetation introduces additional roughness elements, causing an increase in flow resistance as water interacts with the vegetation. This increased resistance leads to higher BSS along the floodplains compared with the relatively smoother main channel.

CONCLUSIONS

The present research investigated the flow dynamics of a compound channel with converging floodplains, focusing on the impact of relative depth and flexible vegetation on flow parameters such as averaged depth velocity and BSS. The key findings of this study are as follows:

1. Converging floodplains cause a significant increase in the DAV because of channel convergence and momentum transfer which results in more concentrated flows.
2. The maximum DAV is significantly dependent on relative depth. As the relative depth increases, the DAV also increases.
3. The flexible vegetation in the floodplains causes resistance to flow which results in a significant drop in the DAV in the entire cross-section.
4. The reduced resistance of flexible vegetation at higher relative depths causes a lateral shift of the zone with maximum DAV towards the floodplains.
5. The BSS increases along the converging reach of the channel due to the convergence of the channel and momentum transfer.
6. In the presence of flexible vegetation, the BSS distribution remains similar to scenarios without vegetation. However, the introduction of flexible vegetation increases flow resistance due to additional roughness elements, resulting in higher BSS along floodplains compared with the smoother main channel.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude for the support provided by the Department of Civil Engineering at Delhi Technological University, Delhi, India. The authors would also like to thank all the staff members in the hydraulic laboratory at the Delhi Technological University.

AUTHOR CONTRIBUTIONS

All authors have read and agreed to the published version of the manuscript.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

FUNDING

This research did not receive any funding.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

Alam, J. & Muzzammil, M. 2011 Flood disaster preparedness in Indian scenario. *International Journal on Recent Trends Engineering & Technology* **5** (3), 34–38. doi:01.IJRTET.05.03.201.

Azamathulla, H. M. & Zahiri, A. 2012 Flow discharge prediction in compound channels using linear genetic programming. *Journal of Hydrology* **454**, 203–207. doi:10.1016/j.jhydrol.2012.05.065.

Bousmar, D., Wilkin, N., Jacquemart, J. H. & Zech, Y. 2004 Overbank flow in symmetrically narrowing floodplains. *Journal of Hydraulic Engineering* **130** (4), 305–312. doi:10.1061/(ASCE)0733-9429(2004)130:4(305).

Brito, M., Fernandes, J. & Leal, J. B. 2016 Porous media approach for RANS simulation of compound open-channel flows with submerged vegetated floodplains. *Environmental Fluid Mechanics* **16**, 1247–1266. doi:10.1007/s10652-016-9481-0.

Das, B. S. & Khatua, K. K. 2018 Flow resistance in a compound channel with diverging and converging floodplains. *Journal of Hydraulic Engineering* **144** (8), 04018051. doi:10.1061/(asce)hy.1943-7900.0001496.

Huai, W., Wang, W. & Zeng, Y. 2013 Two-layer model for open channel flow with submerged flexible vegetation. *Journal of Hydraulic Research* **51** (6), 708–718. doi:10.1080/00221686.2013.818585.

Khatua, K. K., Patra, K. C. & Mohanty, P. K. 2012 Stage-discharge prediction for straight and smooth compound channels with wide floodplains. *Journal of Hydraulic Engineering* **138** (1), 93–99. doi:10.1061/(asce)hy.1943-7900.0000491.

Koftis, T. & Prinos, P. 2018 Reynolds stress modelling of flow in compound channels with vegetated floodplains. *Journal of Applied Water Engineering and Research* **6** (1), 17–27. doi:10.1080/23249676.2016.1209437.

Liu, C., Shan, Y., Liu, X., Yang, K. & Liao, H. 2016 The effect of floodplain grass on the flow characteristics of meandering compound channels. *Journal of Hydrology* **542**, 1–17. doi:10.1016/j.jhydrol.2016.07.037.

Mehrabani, V. F., Mohammadi, M., Ayyoubzadeh, S. A., Fernandes, J. N. & Ferreira, R. M. 2020 Turbulent flow structure in a vegetated non-prismatic compound channel. *River Research and Applications* **36** (9), 1868–1878. doi:10.1002/rra.3723.

Naghavi, M., Mohammadi, M., Mahtabi, G. & Abraham, J. 2023 Experimental assessment of velocity and bed shear stress in the main channel of a meandering compound channel with one-sided blocks in floodplain. *Journal of Hydrology* **617**, 129073. doi:10.1016/j.jhydrol.2023.129073.

Naik, B., Padhi, E. & Khatua, K. K. 2018 Flow prediction of boundary shear stress and depth average velocity of a compound channel with narrowing floodplain. *Iranian Journal of Science and Technology, Transactions of Civil Engineering* **42**, 415–425. doi:10.1007/s40996-018-0105-4.

Nezu, I. & Sanjou, M. 2008 Turburence structure and coherent motion in vegetated canopy open-channel flows. *Journal of Hydro-Environment Research* **2** (2), 62–90. doi:10.1016/j.jher.2008.05.003.

Pandey, M., Azamathulla, H. & Pu, J. H. 2022 *River Dynamics and Flood Hazards: Studies on Risk and Mitigation*. Springer Nature, Singapore. doi:10.1007/978-981-19-7100-6.

Patel, V. C. 1965 Calibration of the Preston tube and limitations on its use in pressure gradients. *Journal of Fluid Mechanics* **23** (1), 185–208. doi:10.1017/S0022112065001301.

Preston, J. 1954 The determination of turbulent skin friction by means of Pitot tubes. *The Aeronautical Journal* **58** (518), 109–121. doi:10.1017/S0368393100097704.

Proust, S., Riviere, N., Bousmar, D., Paquier, A., Zech, Y. & Morel, R. 2006 Flow in compound channel with abrupt floodplain contraction. *Journal of Hydraulic Engineering* **132** (9), 958–970. doi:10.1061/(ASCE)0733-9429(2006)132:9(958).

Proust, S., Fernandes, J. N., Peltier, Y., Leal, J. B., Riviere, N. & Cardoso, A. H. 2013 Turbulent non-uniform flows in straight compound open-channels. *Journal of Hydraulic Research* **51** (6), 656–667. doi:10.1080/00221686.2013.818586.

Rahim, A. S., Yonesi, H. A., Shahinejad, B., Podeh, H. T. & Azamattulla, H. M. 2022 Flow structures in asymmetric compound channels with emergent vegetation on divergent floodplain. *Acta Geophysica* **70** (5), 2403–2421. doi:10.1007/s11600-022-00764-0.

Rahim, A. S., Yonesi, H. A., Rahimi, H. R., Shahinejad, B., Podeh, H. T. & Azamathulla, H. M. 2023 Effect of vegetation on flow hydraulics in compound open channels with non-prismatic floodplains. *AQUA – Water Infrastructure, Ecosystems and Society* **72** (5), 781–797. doi:10.2166/aqua.2023.043.

Rao, P. L., Prasad, B. S. S., Sharma, A. & Khatua, K. K. 2022 Experimental and numerical analysis of velocity distribution in a compound meandering channel with double layered rigid vegetated flood plains. *Flow Measurement and Instrumentation* **83**, 102111. doi:10.1016/j.flowmeasinst.2021.102111.

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Uncorrected Proof

Rezaei, B. 2006 *Overbank Flow in Compound Channels with Prismatic and Non-Prismatic Floodplains*. PhD Thesis, University of Birmingham, Birmingham, UK.

Rezaei, B. & Knight, D. W. 2011 Overbank flow in compound channels with nonprismatic floodplains. *Journal of Hydraulic Engineering* **137** (8), 815–824. doi:10.1061/(asce)hy.1943-7900.0000368.

Uotani, T., Kanda, K. & Michioku, K. 2014 Experimental and numerical study on hydrodynamics of riparian vegetation. *Journal of Hydrodynamics* **26** (5), 796–806. doi:10.1016/s1001-6058(14)60088-3.

Yang, K., Cao, S. & Knight, D. W. 2007 Flow patterns in compound channels with vegetated floodplains. *Journal of Hydraulic Engineering* **133** (2), 148–159. doi:10.1061/(ASCE)0733-9429(2007)133:2(148).

Yonesi, H. A., Omid, M. H. & Ayyoubzadeh, S. A. 2013 The hydraulics of flow in non-prismatic compound channels. *Journal of Civil Engineering and Urbanism* **3** (6), 342–356.

Zhang, J. & Hu, R. 2023 Turbulence structure in an experimental compound channel with varying coverage of riparian vegetation on the floodplain. *Journal of Hydrology* **620**, 129378. doi:10.1016/j.jhydrol.2023.129378.

First received 9 February 2024; accepted in revised form 14 July 2024. Available online 26 July 2024

Downloaded from http://iwaponline.com/ws/article-pdf/doi/10.2166/ws.2024.172/1453689/ws2024172.pdf
by DELHI TECHNOLOGICAL UNIVERSITY (DTU) user

Global Social Welfare
https://doi.org/10.1007/s40609-024-00343-9

Impact of Health Insurance on Accessibility, Utilization of Inpatient Care, and Financial Risk Protection Across States/Union Territories of India

Aashima[1,3] · **Mehak Nanda**[1] · **Rajesh Sharma**[2]

Accepted: 19 June 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Background Several health insurance programmes have been launched in India to improve accessibility to healthcare services and safeguard people against health expenditure catastrophes.

Objective We studied the impact of health insurance enrolment towards accessibility, utilization of inpatient care, and safeguarding against financial risk across all states/union territories (UTs) of India.

Data and Methodology Data from the most recent round (75th round) of the nationally representative survey on health and morbidity, entitled "Social Consumption: Health," was employed in the study. The propensity score matching technique was used in the study to examine the accessibility and utilization of inpatient care, and financial risk protection in the context of health insurance across states/UTs of India.

Results The accessibility to inpatient care was statistically significantly higher among insured than uninsured in 21/32 states/UTs ($p < 0.05$). Hospitalization episodes more than once were also statistically significantly higher among insured than uninsured ($p < 0.05$). However, utilization pattern of inpatient care in terms of incidence of hospitalization with private healthcare providers and longer duration of hospital stay were not impacted much by the health insurance status. Furthermore, health insurance was statistically significantly effective in reducing both out-of-pocket health expenditure (OOPE) and catastrophic health expenditure (CHE) (at 10% threshold) for insured in a few states, namely, Assam, Uttar Pradesh, Karnataka, Maharashtra, Delhi, Punjab, Tamil Nadu, Kerala, Meghalaya, Andhra Pradesh, and Telangana ($p < 0.05$). On the other hand, in some states with good health insurance enrolment rates, including Rajasthan, Chandigarh, Goa, and Chhattisgarh, health insurance did not significantly reduce the OOPE and CHE incidence among insured in comparison to uninsured ($p > 0.05$).

Policy Implications Concerted efforts, such as increasing health insurance enrolment, raising awareness about the existing schemes and their eligibility criteria, and informing beneficiaries about the procedures to avail benefits, are required. There is also an urgent need to include outpatient services under the purview of health insurance to reduce the financial burden and make health insurance more viable. Furthermore, strengthening public healthcare facilities and regulating private healthcare providers are imperative in augmenting financial risk protection in India.

Keywords Health insurance · Inpatient care · Catastrophic health expenditure · Out-of-pocket health expenditure

Introduction

The reliance on out-of-pocket expenditure (OOPE) to finance healthcare is a common feature in many low and middle-income countries (WHO, 2014; Wagstaff et al., 2020). People lacking sufficient financial safeguards face the risk of incurring large unforeseen medical expenditures, which sometimes lead to a reduction in living standards, poverty, and indebtedness (WHO & World Bank, 2021; Rahman et al., 2022). Notably, OOPE pushes nearly 100 million people into poverty worldwide each year (WHO & World Bank, 2017).

✉ Rajesh Sharma
rajesh.sharma@nitkkr.ac.in

[1] University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India

[2] Department of Humanities and Social Sciences, National Institute of Technology Kurukshetra, Kurukshetra, Haryana 136119, India

[3] Jaipuria Institute of Management, Noida, Uttar Pradesh, India

Published online: 12 July 2024

In India, OOPE accounts for 50.6% of the total health expenditure (WHO, 2020a). Several health insurance programmes have been launched in India to improve accessibility to healthcare and safeguard people against associated financial catastrophes (Hooda, 2020). The government of India has initiated social health insurance programmes, such as Employees' State Health Insurance Scheme, 1952 (ESIS), and Central Government Health Scheme, 1954 (CGHS), to provide coverage to formal sector employees. Private health insurance is also available in India, which enables individuals to buy health insurance on their own based on their ability-to-pay premium (Hooda, 2020). Furthermore, government-sponsored health insurance (GSHI) schemes have received traction in India as a measure of risk-pooling and safeguarding against health shocks, mainly for poor and vulnerable populations (Hooda, 2020). Indian constitution has categorized health as a matter of state (Selvaraj et al., 2022); therefore, GSHI schemes are prevalent at both national (e.g., Rashtriya Swasthya Bima Yojsana (RSBY) 2008, Pradhan Mantri Jan Arogya Yojana (PM-JAY) 2018) and state levels (e.g., Rajiv Aarogyasri Health Insurance Scheme (2007) in Andhra Pradesh, Vajpayee Arogyasri Scheme (2009–2010) in Karnataka, Chief Minister's Comprehensive Health Insurance Scheme (2012) in Tamil Nadu) (Hooda, 2020).

In India, there is considerable variation in health insurance enrolment and schemes prevailing across different states (Hooda, 2020; Aashima & Sharma, 2023), highlighting the need to evaluate the impact of health insurance across all states. Previous studies were mainly limited to examining the impact of only GSHI schemes in a few states, namely, Chhattisgarh (Garg et al., 2020; Nandi et al., 2017), Haryana (Gupta et al., 2017; Prinja et al., 2019), Maharashtra (Rao et al., 2014; Rent & Ghosh, 2015; Sabharwal et al., 2014), Gujarat (Devadasan et al., 2013; Gupta et al., 2017; Prinja et al., 2019), Uttar Pradesh (Gupta et al., 2017; Prinja et al., 2019; Raza et al., 2016; Sabharwal et al., 2014), Andhra Pradesh (Garg et al., 2019; Rao et al., 2014), Karnataka (Aggarwal, 2010; Barnes et al., 2017; Garg et al., 2019; Sood et al., 2014), Bihar (Raza et al., 2016), Tamil Nadu (Garg et al., 2019), Kerela (Philip et al., 2016), Jharkhand (Sinha, 2018), and Odisha (Panda & Rout, 2018). A few studies analysed the impact of any health insurance scheme (Apyayee et al., 2021; Gupta, 2021; Mahapatro et al., 2018; Sahoo & Madheswaran, 2014) or any GSHI schemes at the national level only (Ranjan et al., 2018; Sriram & Khan, 2020). Therefore, a comprehensive examination of the impact of health insurance across all Indian states is missing. Also, previous studies were conducted in different years with different datasets, which limits the comparability between them. Furthermore, there is a dearth of studies examining the utilization pattern of inpatient care in the context of health insurance (Prinja et al., 2019; Garg et al., 2019, 2020; Sriram & Khan, 2020). Additionally, previous studies were mainly focused on specific cases (i.e., a particular state or insurance scheme) and lacked in providing a comprehensive examination of the impact of overall health insurance enrolment on accessibility, utilization pattern of inpatient care, and financial risk protection across all states and union territories (UTs) in India.

Against this backdrop, in this study, we first analysed the impact of health insurance in providing access to inpatient care. Second, we analysed the role of health insurance towards the utilization pattern of inpatient care (measured in terms of type of provider, duration of hospital stay, and number of times hospitalized). Third, we assessed the impact of health insurance in safeguarding against financial risk (OOPE, catastrophic health expenditure (CHE) at various thresholds, and impoverishment due to OOPE). The entire analysis was performed separately for each state/UT of India. Furthermore, we review and extensively discuss the findings of the previous literature that examined the impact of the health insurance schemes in providing financial protection in India. This comprehensive assessment of the impact of health insurance across all states/UTs is expected to serve as an informative guide to assess whether health insurance has met its objectives and to identify the states/UTs that require a greater policy attention.

Data and Methodology

Overview of Data Source

The most recent round (75th round) of the nationally representative survey on health and morbidity, entitled "Social Consumption: Health," was employed in the study. The survey was conducted by the National Sample Survey Organisation (NSSO) from July 2017 to June 2018, and covered 555,115 persons (325,883 in rural areas and 229,232 in urban areas) from 113,823 households (64,552 in rural areas and 49,271 in urban areas) in India. The survey collected extensive information about hospitalization, cost of care, nature of ailment, maternal and elderly health dimensions, and demographic and socio-economic characteristics of households and their members. A total of 91,445 hospitalization incidents were reported during the last 365 days. Information regarding the coverage of households under any health insurance scheme was also recorded and categorized into five groups: (i) government-sponsored (e.g., RSBY, Arogyasri), (ii) government employer (e.g., CGHS), (iii) employer-supported (other than government employer, for instance, ESIS), (iv) insurance arranged on their own (also known as private health insurance), and (v) others. Details about the prevalent health insurance programmes are elucidated in Supplementary File (pp 1).

Outcome Variables

The primary outcome variables analysed in the study were accessibility to inpatient care, utilization pattern of inpatient care, and financial risk due to inpatient care in the context

of health insurance. Access to inpatient care is determined by whether an individual has undergone hospitalization during the last 365 days preceding the survey. The utilization pattern of inpatient care was analysed in terms of the type of healthcare provider used (public or private), number of times hospitalized (one or more than once hospitalized), and the duration of hospital stay (short or long duration of hospital stay). The duration of hospital stay was considered "short" if it was up to 6 days and "long" if it exceeded 6 days. Six days was taken as the base because the mean duration of hospital stay in our data was recorded as 5.78 days. The financial protection was measured through OOPE, the incidence of CHE, and impoverishment. We computed OOPE by summing all hospitalization-related expenditures, including package component, doctors'/surgeons' fees, medicines, diagnostic tests, bed charges, other medical expenses, and transportation expenses, and then deducting any reimbursement received. The methodology provided by Wagstaff and Doorslaer (2003) was applied to estimate CHE incidence. The incidence of health expenditure was considered to be catastrophic if OOPE exceeded 10% (CHE-10), 25% (CHE-25), or 40% (CHE-40) of the total annual household consumption expenditure (Garg et al., 2019, 2020, 2022; Ranjan et al., 2018). Furthermore, impoverishment was said to be experienced by individuals whose annual per capita consumption expenditure was initially above the poverty line but later fell below it due to hospitalization-related OOPE (Sangar et al., 2019, 2022). For all outcomes, the impact of overall health insurance enrolment (i.e., enrolled in any health insurance programme) was analysed separately in each state/UT. We restricted our analysis to hospitalization cases only, as most health insurance schemes in India mainly cover inpatient care (La Forgia & Nagpal, 2012; Maurya, 2019).

Statistical Analysis

Propensity Score Matching

The study aims to compare the outcomes between insured and uninsured. However, a direct comparison of outcomes between the groups may lead to a biased estimate of the treatment effect (i.e., health insurance in our case) (Ali et al., 2019; Garrido et al., 2014; Zhao et al., 2021), because treatment groups may not be comparable in their pre-treatment characteristics (Ali et al., 2019; Garrido et al., 2014; Zhao et al., 2021). Therefore, in line with previous studies (Garg et al., 2020; Hou et al., 2022; John et al., 2024; Ranjan et al., 2018; Sabharwal et al., 2014; Sriram & Khan, 2020), we have employed the PSM technique to control for the confounding bias, thereby clearly demonstrating the treatment impact. We estimated the impact of health insurance on accessibility, utilization of inpatient care, and financial risk protection across states/UTs.

We estimated the average treatment effect on the treated (ATET), which measures the average difference in outcomes that the treatment group has obtained with health insurance coverage and the treatment group would have obtained without health insurance. ATET is measured through the following equation.

$$ATET = E(Y_1 - Y_0|(D = 1)) = E(Y_1|D = 1) - E(Y_0|D = 1)$$

In the above equation, ATET represents the change in outcome due to health insurance, Y_1 is the estimate of an outcome if treated (insured), Y_0 is the estimate of an outcome if not treated (not insured), $D=1$ is the participation status in the case of treatment, and $D=0$ is the participation status of the untreated. Variables, namely, age, gender, education, social class, sector (rural/urban), and economic quintiles, were used as matching variables for accessibility and utilization outcomes. Additionally, the type of healthcare provider and ailment category were also used as matching variables for financial protection outcomes. The model was run separately for each state/UTs for each outcome. All the data analysis was carried out using Stata software (version 14.1) (StataCorp, 2015).

Results

Health Insurance Enrolment Across States/UTs

In India, dismally low proportion of the population (15.53%) was enrolled under any health insurance programme, with a substantial variation across states/UTs (Fig. 1). In 22 out of 36 states/UTs, insurance enrolment was lower than the national average. Notably, in Mizoram (78.57%), Andhra Pradesh (72.79%), Chhattisgarh (64.29%), Telangana (61.09%), Dadra and Nagar Haveli (58.25%), and Meghalaya (53.9%), vast majority of the population was enrolled in any health insurance programme. By contrast, in Puducherry, Jammu and Kashmir, Sikkim, Manipur, Uttar Pradesh, Madhya Pradesh, Jharkhand, and Bihar, < 5% of the population was enrolled under any health insurance programme. Of all the health insurance programmes existing in India, the highest enrolment was observed under GSHI schemes (11.70%), followed by CGHS (1.37%) and private health insurance (1.28%) (Supplementary Table 1). In many states/UTs, enrolment was higher in GSHI schemes compared to any other health insurance, while in a few well-off states/UTs, such as Chandigarh (9.75%), Delhi (8.29%), Maharashtra (3.45%), and Haryana (2.42%), higher enrolment was reported under private health insurance. Nearly 5–15% of the population was enrolled under CGHS in a few states/UTs, such as Chandigarh, Mizoram, Lakshadweep, Andaman and Nicobar Islands, Delhi, Daman and Diu, Himachal Pradesh, and Meghalaya. However, enrolment

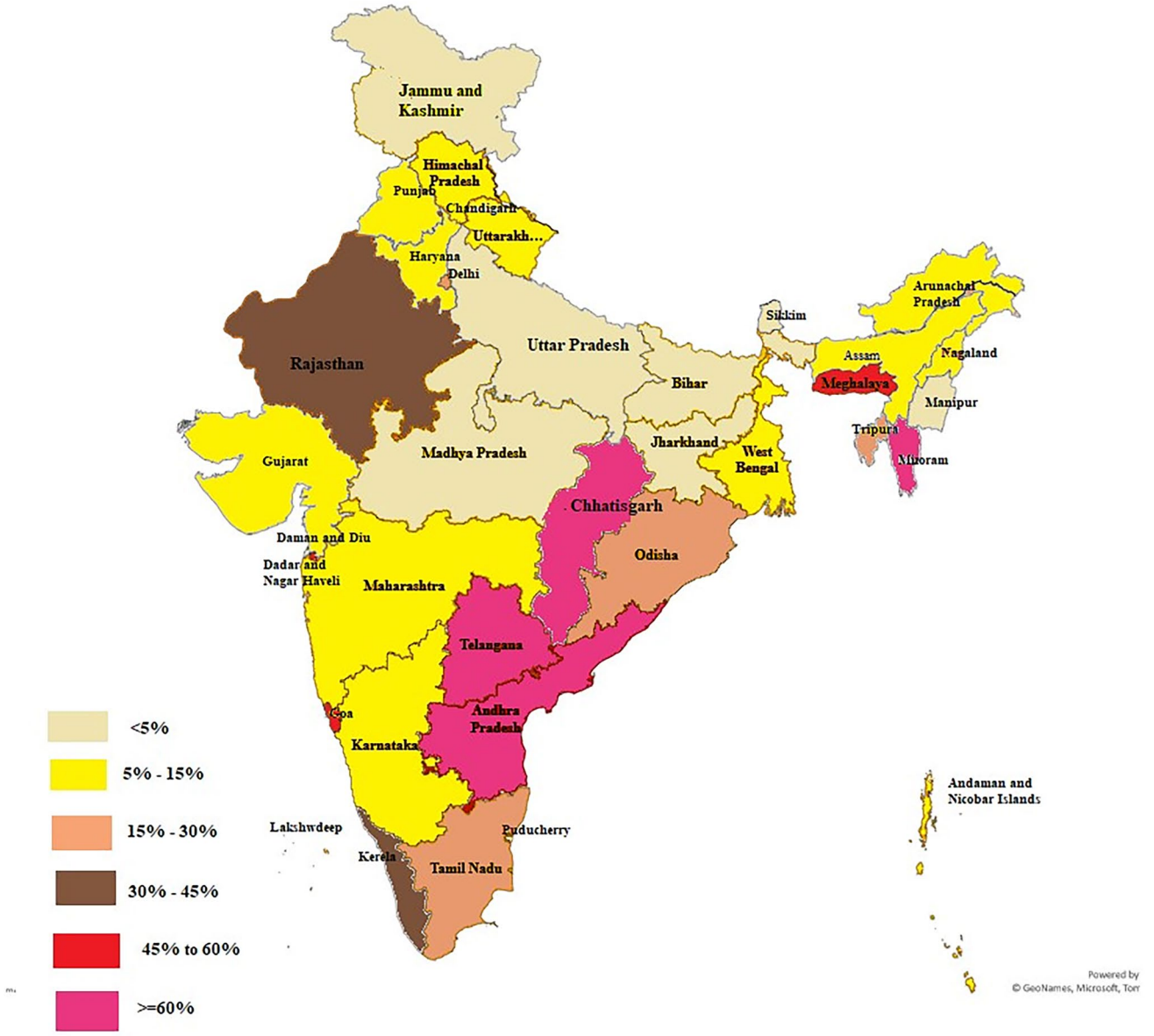


Fig. 1 Health insurance enrolment across states/UTs in India

under ESIS and other health insurance schemes was notably low across most states/UTs.

Impact of Health Insurance on Accessibility and Utilization Pattern of Inpatient Care Across States/UTs

Table 1 presents the impact of health insurance on accessibility and utilization pattern of hospitalization across states/UTs using PSM. Access to inpatient care was statistically significantly higher among insured than uninsured in 21/32 states/UTs ($p < 0.05$). In states/UTs, namely, Chandigarh, Uttarakhand, Tripura, Assam, Chhattisgarh, Dadar and Nagar Haveli, Andhra Pradesh, Goa, Tamil Nadu, Jammu and Kashmir, and Telangana, health insurance enrolment did not result in statistically significantly higher usage of inpatient care among insured than uninsured ($p > 0.05$).

Furthermore, among those who utilized the inpatient care, hospitalization episodes more than once were also statistically significantly higher among insured than uninsured across most states/UTs ($p < 0.05$). On the other hand, hospitalization episodes with private provider were marginally but statistically significantly higher among insured than uninsured in a few states/UTs, namely, Delhi, Karnataka, Maharashtra, West

Table 1 Impact of health insurance on accessibility and utilization of inpatient care across states/UTs

States/UTs	Access to inpatient care	Utilization pattern of inpatient care		
		Private provider	Duration of longer stay	More than once hospitalized
	ATET	ATET	ATET	ATET
Assam	0.0219 (−0.0054 to 0.0491)	0.0467 (−0.0457 to 0.1391)	0.033 (−0.0533 to 0.1194)	0.0348 (−0.0073 to 0.0769)
Bihar	0.1055* (0.0465 to 0.1645)	−0.1526 (−0.2657 to −0.0395)	0.031 (−0.0672 to 0.1291)	−0.0531 (−0.0973 to −0.009)
Jharkhand	0.1322* (0.0576 to 0.2068)	−0.1418 (−0.3244 to 0.0407)	−0.0273 (−0.1732 to 0.1186)	0.0215* (0.0709 to 0.3588)
Madhya Pradesh	0.0366* (0.0081 to 0.0651)	0.0208 (−0.0623 to 0.1038)	0.0739 (−0.0071 to 0.1548)	0.113* (0.0482 to 0.1777)
Odisha	0.0172* (0.0014 to 0.0330)	0.0498* (0.0064 to 0.0931)	0.0553* (0.0079 to 0.1028)	0.0996* (0.0641 to 0.1351)
Uttar Pradesh	0.0333* (0.0109 to 0.0557)	−0.0066 (−0.0611 to 0.0478)	0.0672* (0.0044 to 0.1300)	0.1002* (0.0507 to 0.1498)
Arunachal Pradesh	0.0734* (0.0453 to 0.1016)	0.0448 (−0.0281 to 0.1177)	−0.0056 (−0.1022 to 0.0909)	−0.0256 (−0.0504 to −0.0007)
Gujarat	0.0246* (0.0104 to 0.0387)	−0.0097 (−0.0543 to 0.0348)	0.0328 (−0.0099 to 0.0754)	0.1171* (0.0864 to 0.1479)
Manipur	0.0989* (0.0445 to 0.1533)	0.0105 (−0.0693 to 0.0904)	0.1215* (0.0173 to 0.2257)	−0.0005 (−0.0016 to 0.0005)
Nagaland	0.069* (0.0251 to 0.1129)	−0.0709 (−0.1736 to 0.0318)	−0.0035 (−0.0553 to 0.0482)	−0.0094 (−0.0275 to 0.0087)
Sikkim	0.2409* (0.1583 to 0.3234)	0.0268* (0.1535 to 0.3835)	0.0189* (0.0450 to 0.3347)	0.0357 (−0.0303 to 0.1017)
Tripura	0.0178 (−0.0099 to 0.0455)	0.0371* (0.0002 to 0.0741)	0.0341 (−0.0395 to 0.1076)	0.1091* (0.0569 to 0.1614)
Uttarakhand	0.0175 (−0.0189 to 0.0539)	−0.01 (−0.1236 to 0.1036)	0.053 (−0.0592 to 0.1653)	0.1207* (0.0634 to 0.1779)
Haryana	0.0518* (0.0292 to 0.0745)	0.0521 (−0.0117 to 0.1158)	0.0467 (−0.0143 to 0.1078)	0.0661 (0.0155 to 0.1167)
Jammu and Kashmir	0.0222 (−0.0021 to 0.0466)	−0.0116 (−0.0836 to 0.0603)	−0.0293 (−0.0866 to 0.0279)	−0.002 (−0.0419 to 0.0378)
Karnataka	0.0361* (0.0160 to 0.0563)	0.0634* (0.0239 to 0.1030)	0.0515* (0.0012 to 0.1019)	0.089* (0.0592 to 0.1187)
Maharashtra	0.0425* (0.0279 to 0.0571)	0.0417* (0.0152 to 0.0682)	0.0458* (0.0077 to 0.0839)	0.0765* (0.0471 to 0.1059)
West Bengal	0.0265* (0.0136 to 0.0393)	0.0989* (0.0639 to 0.1339)	0.0202 (−0.0171 to 0.0575)	0.0413* (0.0103 to 0.0723)
Delhi	0.0271* (0.0025 to 0.0518)	0.1941* (0.1157 to 0.2724)	0.0229 (−0.057 to 0.1029)	0.0275 (0 to 0.0550)
Himachal Pradesh	0.0467* (0.0213 to 0.0720)	0.0827 (0.0094 to 0.1561)	0.1138* (0.0449 to 0.1827)	0.1828* (0.1227 to 0.2430)
Punjab	0.0484* (0.0255 to 0.0712)	−0.1206* (−0.1883 to −0.0530)	0.0869* (0.0241 to 0.1496)	0.0622* (0.0169 to 0.1075)
Tamil Nadu	0.0103 (−0.0024 to 0.0230)	0.0293 (−0.0064 to 0.0650)	0.0237 (−0.0133 to 0.0606)	0.0731* (0.0470 to 0.0992)
Rajasthan	0.0097* (0.0008 to 0.0185)	0.0448 (0.009 to 0.0806)	−0.0484* (−0.0786 to −0.0183)	0.0847* (0.062 to 0.1074)
Chandigarh	0.0315 (−0.0208 to 0.0837)	0.0945 (−0.0621 to 0.2511)	0.0219 (−0.1281 to 0.1719)	0.1813* (0.0851 to 0.2774)
Goa	−0.0248 (−0.0669 to 0.0173)	−0.0445 (−0.1804 to 0.0913)	−0.0812 (−0.2357 to 0.0733)	0.0278 (−0.0264 to 0.0820)
Kerala	0.0219* (0.0091 to 0.0346)	−0.1021* (−0.1352 to −0.0689)	0.0215 (−0.0112 to 0.0542)	0.0274 (−0.0005 to 0.0553)
Chhattisgarh	0.0056 (−0.0079 to 0.0190)	0.0088 (−0.0426 to 0.0603)	0.0179 (−0.0256 to 0.0615)	0.0401* (0.0185 to 0.0617)
Meghalaya	0.0416* (0.0252 to 0.0579)	0.0794* (0.0191 to 0.1396)	0.0267 (−0.0397 to 0.0930)	0 (−0.0006 to 0.0006)
Mizoram	0.0678* (0.0462 to 0.0894)	−0.0238 (−0.1220 to 0.0744)	0.0692 (−0.0484 to 0.1869)	−0.0023 (−0.0234 to 0.0188)
Dadra and Nagar Haveli	−0.0661 (−0.1363 to 0.0041)	0.0449 (−0.1633 to 0.253)	−0.0192 (−0.0954 to 0.057)	0.0897* (0.0161 to 0.1634)
Andhra Pradesh	−0.0017 (−0.0168 to 0.0134)	−0.0749 (−0.1178 to −0.032)	0.0082 (−0.0460 to 0.0625)	0.0957* (0.0717 to 0.1196)
Telangana	0.0072 (−0.0096 to 0.0241)	−0.0831 (−0.1343 to −0.0319)	−0.0379 (−0.082 to 0.0061)	0.0302* (0.0151 to 0.0453)

Puducherry, Andaman and Nicobar Islands, Lakshadweep, and Daman and Diu union territories were dropped from the analysis due to the small sample size

ATET average treatment effect on treated

*$p < 0.05$; figures inside square brackets represent 95% confidence interval

Bengal, Sikkim, Tripura, Odisha, and Meghalaya ($p < 0.05$). In Kerala and Punjab, the hospitalization episodes with private provider were statistically significantly lower among insured than uninsured ($p < 0.05$). Similarly, the episodes of longer duration of hospital stay were statistically significantly higher among insured in only a few states, namely, Maharashtra, Himachal Pradesh, Punjab, Odisha, Uttar Pradesh, Manipur, and Sikkim ($p < 0.05$) (Table 1).

Impact of Health Insurance on Financial Risk Protection Across States/UTs

Table 2 presents the impact of health insurance on OOPE, CHE, and impoverishment. Health insurance enrolment statistically significantly reduced the OOPE for insured than uninsured in 12 states/UTs ($p < 0.05$). A substantial decrease in OOPE among insured was observed in Assam, Uttar Pradesh, Karnataka, Maharashtra, Delhi, and Punjab than uninsured ($p < 0.05$). The incidence of CHE-10 was statistically significantly lower among insured by more than 10% in Assam, Uttar Pradesh, Uttarakhand, Jammu and Kashmir, Maharashtra, Delhi, Punjab, Dadra and Nagar Haveli, and Andhra Pradesh ($p < 0.05$). In states, namely, Gujarat, Karnataka, Tamil Nadu, Kerala, Meghalaya, and Telangana, the reduction in CHE-10 among insured was between 5% and 10%. In all other states, health insurance was observed to be statistically ineffective in reducing the financial burden ($p > 0.05$). In some states with good health insurance enrolment rates, such as Rajasthan, Chandigarh, Goa, and Chhattisgarh, health insurance did not statistically significantly reduce the OOPE and CHE incidence among insured than uninsured ($p > 0.05$). Furthermore, CHE-25 was statistically significantly lower among insured in Delhi, Uttar Pradesh, Maharashtra, Andhra Pradesh, Kerala, and Telangana ($p < 0.05$). The CHE-40 incidence was statistically significantly lower in only Andhra Pradesh, Telangana, and Maharashtra ($p < 0.05$). Lastly, health insurance enrolment reduced the impoverishment incidence in a few states such as Punjab, Assam, Maharashtra, Kerala, and Telangana ($p < 0.05$). These findings suggest that health insurance enrolment has alleviated the financial burden among insured to some extent, albeit limited to certain states (Table 2).

Discussion

To the best of our knowledge, this is the first study that analysed the role of health insurance across several dimensions, namely, accessibility to inpatient care, utilization pattern of inpatient care, and financial risk protection across all states/UTs in India.

We observed a small but statistically significantly higher accessibility to inpatient care among insured than uninsured across most states/UTs of India. Previous studies have also found higher accessibility to inpatient care among insured (Mahapatro et al., 2018; Nandi et al., 2017; Prinja et al., 2017; Reshmi et al., 2021; Sriram & Khan, 2020). The hospitalization episodes more than once were also found to be higher among insured than uninsured. This observed trend could be attributed to a genuine decline in financial barriers to accessing care among insured, or due to moral hazard, or provider-induced demand (Dang et al., 2021; Prinja et al., 2017; Sengupta & Rooj, 2019; Sriram & Khan, 2020). Furthermore, we observed that the hospitalization episodes with private provider were only marginally higher among insured than uninsured in few states only. This may be attributed to the high OOPE in private health facilities irrespective of the health insurance status as highlighted by the previous studies (Nandi et al., 2017; Garg et al., 2019, 2020; Ranjan et al., 2018; Muraleedharan et al., 2020; Aashima & Sharma, 2024a). Also, the utilization of the type of provider depends upon the availability (Baru et al., 2010; Rout et al., 2021) and empanelment under different health insurance programmes (Smith et al., 2019). For instance, as per a policy brief, no private hospital was empanelled in aspirational districts[1] of nine states under the recently launched GSHI scheme, PM-JAY (Smith et al., 2019). Additionally, episodes of longer duration of hospital stay were also not statistically different among insured and uninsured across most states/UTs, which is in line with a previous study focused on poor individuals (Sriram & Khan, 2020), which argues that the length of hospital stay is primarily determined by the patient's health condition, and healthcare providers have little incentive to prolong stays due to low package rates in GSHI schemes (Sriram & Khan, 2020). Additionally, providers are often unaware of the patient's insurance status as this is typically handled by hospital administrative divisions (Sriram & Khan, 2020). By contrast, in Maharashtra, Himachal Pradesh, Punjab, Odisha, Uttar Pradesh, Manipur, and Sikkim, the episodes of longer duration of hospital stay were higher among insured than uninsured. Overall, this reflects that health insurance enrolment has increased the accessibility to inpatient care among insured; however, the utilization pattern of inpatient care (i.e., private provider usage and longer duration of hospital stay) was not impacted much by the health insurance status.

We found that health insurance was statistically significantly effective in reducing both OOPE and CHE (at 10% threshold) for insured in some states such as Assam, Uttar Pradesh, Karnataka, Maharashtra, Delhi, Punjab, Tamil Nadu, Kerala, Meghalaya, Andhra Pradesh, and Telangana. However, the incidence of CHE at 25% and 40% threshold and impoverishment were not statistically different among insured and uninsured across most states. Furthermore, in some states with good health insurance enrolment along with state's own GSHI schemes, such as Rajasthan, Goa, and Chhattisgarh, health insurance was not found to be effective in reducing the financial burden for the insured.

Previous literature reported the lack of effectiveness of GSHI schemes in providing the financial protection in states

[1] Aspirational districts are the most under-developed districts across the country, which are chosen to quickly and effectively transform under "Aspirational districts" programme.

Table 2 Impact of health insurance on OOPE, CHE, and impoverishment across states/UTs

State/union territories	OOPE ATET	CHE 10 ATET	CHE 25 ATET	CHE 40 ATET	Impoverishment ATET
Assam	−13,665.78* (−25,433.72 to −1897.85)	−0.1573* (−0.2597 to −0.0549)	−0.0820 (−0.1729 to 0.0090)	−0.0407 (−0.1014 to 0.0199)	−0.0663* (−0.1282 to −0.0044)
Bihar	3907.622 (−27,892.58 to 35,707.83)	−0.0322 (−0.2225 to 0.1582)	−0.0231 (−0.1370 to 0.0908)	−0.0109 (−0.0747 to 0.0529)	−0.0072 (−0.0617 to 0.0473)
Jharkhand	3649.513 (−11,577.82 to 18,876.85)	0.024 (−0.1148 to 0.1628)	0.0625 (−0.0561 to 0.1811)	0.0449 (−0.0655 to 0.1553)	0.0387 (−0.0330 to 0.1103)
Madhya Pradesh	−2189.945 (−14,338.77 to 9958.88)	−0.1108 (−0.2056 to −0.0161)	−0.0147 (−0.0843 to 0.0549)	−0.0056 (−0.0596 to 0.0484)	0.0128 (−0.0258 to 0.0513)
Odisha	619.3699 (−2806.87 to 4045.61)	−0.0105 (−0.0651 to 0.0442)	0.0125 (−0.0330 to 0.0579)	0.0215 (−0.0166 to 0.0597)	−0.014 (−0.0525 to 0.0244)
Uttar Pradesh	−12,610.74* (−22,149.65 to −3071.84)	−0.1385* (−0.2114 to −0.0657)	−0.0693* (−0.1182 to −0.0205)	−0.0231 (−0.0585 to 0.0122)	−0.0177 (−0.0475 to 0.0121)
Arunachal Pradesh	−1048.416 (−3350.50 to 1253.67)	−0.0228 (−0.1101 to 0.0646)	−0.0142 (−0.0538 to 0.0255)	−0.0201 (−0.0449 to 0.0047)	−0.0481 (−0.1075 to 0.0113)
Gujarat	−2051.341 (−6276.99 to 2174.31)	−0.0805* (−0.1276 to −0.0334)	−0.0460* (−0.0840 to −0.0080)	−0.0074 (−0.0379 to 0.0231)	0.0147 (−0.0107 to 0.0402)
Manipur	−877.9216 (−5724.60 to 3968.76)	−0.1013 (−0.2227 to 0.0201)	0.0062 (−0.0704 to 0.0829)	−0.0182 (−0.0505 to 0.0141)	−0.0092 (−0.0831 to 0.0647)
Nagaland	−592.0262 (−6992.05 to 5807.99)	−0.1156 (−0.2087 to −0.0226)	−0.0162 (−0.0515 to 0.0191)	0.0031 (−0.0106 to 0.0168)	0.0012 (−0.0510 to 0.0510)
Sikkim	−3119.727 (−8847.80 to 2608.34)	−0.2317 (−0.3786 to −0.0849)	0.0013 (−0.0475 to 0.0475)	0.0159 (−0.0379 to 0.0696)	−0.0055 (−0.0730 to 0.0621)
Tripura	−255.005 (−3161.77 to 2651.76)	−0.0164 (−0.0816 to 0.0488)	0.0261 (−0.0088 to 0.061)	−0.0003 (−0.0230 to 0.0225)	0.028 (−0.0028 to 0.0588)
Uttarakhand	−6183.068 (−19,460.05 to 7093.92)	−0.1382* (−0.2497 to −0.0267)	−0.0348* (−0.1193 to 0.0498)	0.0354 (−0.0337 to 0.1046)	0.003 (−0.0617 to 0.0677)
Haryana	−2123.616 (−8185.25 to 3938.02)	−0.0474 (−0.1159 to 0.0210)	−0.0059 (−0.0462 to 0.0343)	0.0131 (−0.0175 to 0.0437)	−0.004 (−0.0366 to 0.0287)
Jammu and Kashmir	−4257.871 (−11,335.16 to 2819.41)	−0.117* (−0.1961 to −0.0379)	0.0063 (−0.0425 to 0.0551)	0.0097 (−0.0247 to 0.0440)	−0.0025 (−0.0319 to 0.0269)
Karnataka	−10,302.05* (−14,386.28 to −6217.83)	−0.0935* (−0.1481 to −0.0388)	−0.0374* (−0.0772 to 0.0024)	−0.0126 (−0.0396 to 0.0144)	−0.0023 (−0.0287 to 0.0242)
Maharashtra	−12,041.91* (−18,650.50 to −5433.33)	−0.1822* (−0.2272 to −0.1372)	−0.0566* (−0.0929 to −0.0203)	−0.0247* (−0.0514 to 0.0020)	−0.0386* (−0.0659 to −0.0113)
West Bengal	−5228.565 (−11,194.53 to 737.40)	−0.1259 (−0.1649 to −0.0870)	−0.0305 (−0.0618 to 0.0008)	−0.0049 (−0.0294 to 0.0197)	−0.0273 (−0.052 to −0.0025)
Delhi	−21,098.83* (−35,956.46 to −6241.21)	−0.2033* (−0.2967 to −0.1100)	−0.0784* (−0.1529 to −0.0038)	−0.0412 (−0.0989 to 0.0164)	−0.0084 (−0.0468 to 0.0301)
Himachal Pradesh	3206.737 (−6958.89 to 13,372.37)	−0.0173 (−0.0926 to 0.0580)	0.0431 (−0.0095 to 0.0957)	0.0463 (0.0029 to 0.0898)	0.0080 (−0.0408 to 0.0567)

Table 2 (continued)

State/union territories	OOPE ATET	CHE 10 ATET	CHE 25 ATET	CHE 40 ATET	Impoverishment ATET
Punjab	−17,159.67*	−0.1068*	−0.0470	−0.0341	−0.062*
	(−28,017.46 to −6301.87)	(−0.1761 to −0.0375)	(−0.1025 to 0.0085)	(−0.0806 to 0.0124)	(−0.1156 to −0.0083)
Tamil Nadu	−3963.50*	−0.0493*	−0.0225	−0.0162	−0.0198
	(−7958.38 to 31.48)	(−0.0835 to −0.0152)	(−0.0511 to 0.0061)	(−0.0400 to 0.0077)	(−0.0455 to 0.0060)
Rajasthan	118.1636	0.01	0.0017	0.0047	0.0067
	(−2299.73 to 2536.06)	(−0.0220 to 0.0421)	(−0.0230 to 0.0263)	(−0.0128 to 0.0223)	(−0.0170 to 0.0305)
Chandigarh	−31,341.31	−0.1639	−0.1039	−0.1299	−0.1288
	(−66,546.27 to 3863.65)	(−0.3368 to 0.0090)	(−0.2591 to 0.0513)	(−0.2716 to 0.0118)	(−0.2936 to 0.0360)
Goa	−4551.75	−0.007	−0.0499	−0.0049	−0.0554
	(−11,866.59 to 2763.10)	(−0.0948 to 0.0807)	(−0.1212 to 0.0214)	(−0.0333 to 0.0235)	(−0.1270 to 0.0162)
Kerala	−6596.52*	−0.0686*	−0.0317*	−0.0095	−0.0222*
	(−9785.82 to −3407.21)	(−0.1003 to −0.0370)	(−0.0554 to −0.0079)	(−0.0274 to 0.0084)	(−0.0452 to 0.0008)
Chhattisgarh	−1907.579	−0.0341	−0.0442	−0.0117	−0.0200
	(−19,995.53 to 16,180.37)	(−0.0804 to 0.0122)	(−0.0875 to −0.0010)	(−0.0438 to 0.0204)	(−0.0520 to 0.0121)
Meghalaya	−4595.981*	−0.0509*	−0.0489	−0.0352	−0.0052
	(−8782.34 to −409.62)	(−0.1041 to 0.0023)	(−0.1036 to 0.0057)	(−0.0810 to 0.0105)	(−0.0290 to 0.0185)
Mizoram	−4161.77*	−0.1099	−0.0166	0.0014	0.0051
	(−6068.05 to −2255.49)	(−0.1928 to −0.0270)	(−0.0412 to 0.0081)	(−0.0157 to 0.0184)	(0.0006 to 0.0096)
Dadra and Nagar Haveli	−243.62	−0.2564*	0.0064	0.0128	0.0278
	(−4372.24 to 3884.99)	(−0.4299 to −0.0829)	(−0.0224 to 0.0352)	(−0.0128 to 0.0384)	(−0.0118 to 0.0673)
Andhra Pradesh	−5242.29*	−0.11*	−0.0532*	−0.0486*	−0.0291
	(−9166.62 to −1317.96)	(−0.1587 to −0.0612)	(−0.0980 to −0.0084)	(−0.0898 to −0.0074)	(−0.0706 to 0.0123)
Telangana	−4593.48*	−0.0768*	−0.0582*	−0.0371*	−0.0366*
	(−8682.34 to −504.61)	(−0.1229 to −0.0306)	(−0.0980 to −0.0185)	(−0.0750 to 0.0008)	(−0.0694 to −0.0038)

Puducherry, Andaman and Nicobar Islands, Lakshadweep, and Daman and Diu union territories were dropped from the analysis due to the small sample size

ATET average treatment effect on treated

*$p < 0.05$; figures inside square brackets represent 95% confidence interval

such as Chhattisgarh, Gujarat, Haryana, Uttar Pradesh, Maharashtra, Karnataka, Andhra Pradesh, Tamil Nadu, and Jharkhand (Supplementary Table 2). On the other hand, few studies conducted in Bihar, Haryana, Karnataka, Uttar Pradesh, and Maharashtra reported the effectiveness of health insurance schemes in reducing the financial burden. It was observed that previous studies examining any health insurance scheme or any GSHI scheme at the national level also reported marginal effectiveness of health insurance in providing financial risk protection to the insured (Supplementary Table 2). The difference in results could be attributed to the differences in data, year, methodology, and scheme analysed (as most studies were restricted to analysing state-specific GSHI schemes). The prominent reasons highlighted by the previous literature for the limited effectiveness of GSHI schemes in reducing financial burden were (i) low awareness among beneficiaries regarding various facets of health insurance (such as information regarding stipulated benefits one is entitled to, the procedure to avail benefits, number of family members covered, list of empanelled hospitals, and ailments covered) (Ahlin et al., 2016; Devadasan et al., 2013; Rajasekhar et al., 2011; Rent & Ghosh, 2015; Sabharwal et al., 2014; Thakur, 2016) and (ii) continued spendings on medicines, diagnostics, and consumables (Ahlin et al., 2016; Devadasan et al., 2013;

Ecks, 2021; Kamath & Kamath, 2020; Nandi & Schneider, 2020; Rent & Ghosh, 2015; Sabharwal et al., 2014; Selvaraj & Karan, 2012). Studies have also reported malpractices by empanelled private hospitals, such as overbilling of insurance patients, persuading beneficiaries to utilize services that are not covered, performing unnecessary procedures on insured patients to claim insurance money, and lack of transparency in billing (Chakravarthi, 2018; Dehury et al., 2019; Selvaraj et al., 2022; Kamath & Kamath, 2020; Oxfam India, 2021). The inadequate monitoring of private healthcare providers is also a concerning issue (Chakravarthi, 2018; Dehury et al., 2019; Selvaraj et al., 2022; Oxfam India, 2021). Furthermore, problems with beneficiary card, long waiting time at empanelled hospitals, reimbursement issues, low package rates, and no coverage for outpatient care were other plausible reasons reported for less-than-optimal outcomes of health insurance in India (Ahlin et al., 2016; Devadasan et al., 2013; Kamath & Kamath, 2020; Rajasekhar et al., 2011; Sabharwal et al., 2014; Thakur, 2016).

There is a pressing need to strengthen the public health sector, improve regulations for healthcare providers, and address challenges encountered under health insurance schemes to augment financial risk protection in India. Also, it is important to increase public health expenditure in India. According to a WHO report (2020b), countries with higher public health expenditures can offer greater financial protection against catastrophic and impoverishing health spending. Furthermore, health insurance in India mainly provides coverage for the inpatient care, and outpatient care remains outside the ambit of health insurance. Even the recently launched central-level GSHI scheme, PM-JAY, which aims to cover the bottom 40% of the Indian population, does not provide coverage for outpatient services (Mukherjee & Chaudhuri, 2020), which puts a colossal financial burden of accessing healthcare in India (Gupta et al., 2016; Sangar et al., 2022; Nanda & Sharma, 2023a; Aashima & Sharma, 2024b). India is undergoing an epidemiological transition with an increasing burden of non-communicable diseases (Patel et al., 2011; Yadav & Arokiasamy, 2014), which requires frequent consultations, regular visits to a doctor, diagnostic tests, and long-term medication support to manage the disease progression (Nanda & Sharma, 2023b). Covering outpatient services can improve access to primary care and promote the viability of inpatient health insurance schemes (Taneja & Taneja, 2016). Evidence suggests that the composition of the benefit package has an impact on utilization and financial risk protection (Watson et al., 2021). For instance, in China, the New Rural Cooperative Medical System was reformed to include outpatient services, and as a result, the healthcare utilization by the poor has increased in the country (Watson et al., 2021).

Strengths and Limitations

The study leveraged the recent NSSO 2018 survey data, offering a notable improvement over earlier studies that were based upon previous NSSO rounds. Also, the study meticulously integrates insights from a wide array of existing literature. Furthermore, our study significantly contributes to the existing body of research by providing a comprehensive overview of the impact of health insurance on accessibility, utilization pattern of inpatient care, and financial risk across all states/UTs of India.

As with all studies, this study also has a few limitations. Survey information regarding hospitalization episodes involves a long recall period of 365 days, making it prone to recall bias (Yadav et al., 2024). The study relies on a cross-sectional survey that limits its use for causal inference and analyses of the extent of the economic strain and the disruption to living standards caused by health expenditures in the long term. Last, through this study, we couldn't estimate the impact of the recently launched PM-JAY scheme, which could be gauged in the coming years through future rounds of NSSO.

Conclusion and Policy Recommendations

We found that health insurance enrolment has increased accessibility to inpatient care among insured; however, the utilization pattern of inpatient care in terms of utilizing private healthcare provider and longer duration of hospital stay was not impacted much by the health insurance status across states/UTs. Also, health insurance was observed to be effective in providing financial protection to the insured in some states. Concerted efforts, such as increasing health insurance enrolment, raising awareness about prevalent schemes and their eligibility criteria, and informing beneficiaries about the procedures to avail benefits, are required. There is also an urgent need to include outpatient services under the purview of health insurance to reduce the financial burden and make health insurance more viable. Better monitoring, supervision, and regulation by state authorities are required to provide effective coverage under health insurance. Furthermore, strengthening public healthcare facilities, regulating private healthcare providers, and incentivizing the public healthcare sector to compete for better service provision are equally important in delivering quality and affordable healthcare services in India.

Abbreviations WHO: World Health Organization; OOPE: Out-of-pocket expenditure; ESIS: Employees' State Insurance Scheme; CGHS: Central Government Health Scheme; GSHI: Government-sponsored health insurance; RSBY: Rashtriya Swasthya Bima Yojana; PM-JAY: Pradhan Mantri Jan Aarogya Yojana; UTs: Union territories; CHE: Catastrophic health expenditure; NSSO: National Sample Survey

Office; PSM: Propensity score matching; ATET: Average treatment effect on the treated

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40609-024-00343-9.

Data Availability Data used in the manuscript is available in the public domain. Data can be assessed from https://microdata.gov.in/nada43/index.php/catalog/152.

Declarations

Ethics Approval and Consent to Participate The study is based on secondary data available in the public domain, hence, does not require any ethical approval.

Competing Interests The authors declare no competing interests.

References

Aashima, & Sharma, R. (2023). Inequality and disparities in health insurance enrolment in India. *Journal of Medicine Surgery and Public Health, 1*, 100009.

Aashima, & Sharma, R. (2024a). Is health insurance really benefitting Indian population? Evidence from a nationally representative sample survey. *The International Journal of Health Planning and Management, 39*(2), 293–310.

Aashima, & Sharma, R. (2024b). An examination of inter-state variation in utilization of healthcare services, associated financial burden and inequality: Evidence from nationally representative survey in India. *International Journal of Social Determinants of Health and Health Services*. https://doi.org/10.1177/27551938241230761

Aggarwal, A. (2010). Impact evaluation of India's 'Yeshasvini' community-based health insurance programme. *Health Economics, 19*(S1), 5–35.

Ahlin, T., Nichter, M., & Pillai, G. (2016). Health insurance in India: What do we know and why is ethnographic research needed. *Anthropology & Medicine, 23*(1), 102–124.

Ali, M. S., Prieto-Alhambra, D., Lopes, L. C., Ramos, D., Bispo, N., Ichihara, M. Y., ... Smeeth, L. (2019). Propensity score methods in health technology assessment: Principles, extended applications, and recent advances. *Frontiers in pharmacology, 10*, 973.

Apyayee, S., Sourav, D., & Mithun, M. (2021). How does health insurance affect healthcare spending in India over the decade: A two part model approach. *Epidemiol Sci, 11*(401), 2.

Barnes, K., Mukherji, A., Mullen, P., & Sood, N. (2017). Financial risk protection from social health insurance. *Journal of Health Economics, 55*, 14–29.

Baru, R., Acharya, A., Acharya, S., Kumar, A. S., & Nagaraj, K. (2010). Inequities in access to health services in India: Caste, class and region. *Economic and political Weekly, 45*(38), 49–58.

Chakravarthi, I. (2018). Regulation of private health care providers in India: Current status, future directions. *Indian Journal of Public Administration, 64*(4), 587–598.

Dang, A., Dang, D., & Vallish, B. N. (2021). Importance of evidence-based health insurance reimbursement and health technology assessment for achieving universal health coverage and improved access to health in India. *Value in Health Regional Issues, 24*, 24–30.

Dehury, R. K., Samal, J., Coutinho, S., & Dehury, P. (2019). How does the largely unregulated private health sector impact the Indian mass? *Journal of Health Management, 21*(3), 383–393. https://doi.org/10.1177/0972063419868561

Devadasan, N., Seshadri, T., Trivedi, M., & Criel, B. (2013). Promoting universal financial protection: Evidence from the Rashtriya Swasthya Bima Yojana (RSBY) in Gujarat, India. *Health Research Policy and Systems, 11*(1), 1–8.

Ecks, S. (2021). "Demand side" health insurance in India: The price of obfuscation. *Medical Anthropology, 40*(5), 404–416.

Garg, S., Bebarta, K. K., & Tripathi, N. (2020). Performance of India's national publicly funded health insurance scheme, Pradhan Mantri Jan Arogaya Yojana (PMJAY), in improving access and financial protection for hospital care: Findings from household surveys in Chhattisgarh state. *BMC Public Health, 20*(1), 1–10.

Garg, S., Chowdhury, S., & Sundararaman, T. (2019). Utilisation and financial protection for hospital care under publicly funded health insurance in three states in Southern India. *BMC Health Services Research, 19*(1), 1–11.

Garg, S., Bebarta, K. K., & Tripathi, N. (2022). Role of publicly funded health insurance in financial protection of the elderly from hospitalisation expenditure in India-findings from the longitudinal aging study. *BMC Geriatrics, 22*(1), 572.

Garrido, M. M., Kelley, A. S., Paris, J., et al. (2014). Methods for constructing and assessing propensity scores. *Health Services Research, 49*(5), 1701–1720. https://doi.org/10.1111/1475-6773.12182

Gupta, I., Chowdhury, S., Trivedi, M., & Prinja, S. (2017). Do health coverage schemes ensure financial protection from hospitalization expenses? Evidence from eight districts in India. *Journal of Social and Economic Development, 19*(1), 83–93.

Gupta, I., Chowdhury, S., Prinja, S., & Trivedi, M. (2016). Out-of-pocket spending on out-patient care in India: Assessment and options based on results from a district level survey. *PLoS ONE, 11*(11), e0166775.

Gupta, V. (2021). Health care in India: Evaluation of health insurance in India using NSS data. सर्वेक्षण, 14. *Journal of National Statistical Office*. Available: https://mospi.gov.in/sites/default/files/publication_reports/110th_111th_Sarvekshana.pdf#page=21

Hooda, S. K. (2020). Penetration and coverage of government-funded health insurance schemes in India. *Clinical Epidemiology and Global Health, 8*(4), 1017–1033. https://doi.org/10.1016/j.cegh.2020.03.014

Hou, B., Wu, Y., & Huang, S. (2022). Participating in health insurance and health improvements for the relatively poor population: A propensity score analysis. *Frontiers in Public Health, 10*, 968009.

John, D., Yadav, J., Ray, D., Bhattacharya, P., Mukherjee, N., Patil, R., ... Hansda, M. (2024). Hospitalisation expenditure on tuberculosis among tribal populations in India: A repeated cross-sectional analysis of national sample survey data, 2004 to 2018. *Public Health in Practice, 7*, 100490. https://doi.org/10.1016/j.puhip.2024.100490

Kamath, S., & Kamath, R. (2020). Shortcomings in India's first national attempt at universal healthcare through publicly funded health insurance. *Postgraduate Medical Journal, 96*(1138), 449–450.

La Forgia, G., & Nagpal, S. (2012). *Government-sponsored health insurance in India: Are you covered?* World Bank Publications.

Mahapatro, S. R., Singh, P., & Singh, Y. (2018). How effective health insurance schemes are in tackling economic burden of healthcare in India. *Clinical Epidemiology and Global Health, 6*(2), 75–82. https://doi.org/10.1016/j.cegh.2017.04.002

Maurya, D. (2019). Understanding public health insurance in India: A design perspective. *Int J Health Plan Manag, 34*(4), e1633–e1650. https://doi.org/10.1002/hpm.2856

Mukherjee, S., & Chaudhuri, A. (2020). Outpatient care and expenses: Can they be ignored in health insurance programmes. *Institute of Development Studies Kolkata*. Available: https://idsk.edu.in/wp-content/uploads/2020/06/OP-68.pdf

Muraleedharan, V. R., Vaidyanathan, G., Sundararaman, T., Dash, U., Ranjan, A., & Rajesh, M. (2020). Invest more in public healthcare facilities. *Econ Polit Wkly, 55*(37), 53–60.

Nanda, M., & Sharma, R. (2023a). A comprehensive examination of the economic impact of out-of-pocket health expenditures in India. *Health Policy and Planning, 38*(8), 926–938.

Nanda, M., & Sharma, R. (2023b). Financial burden of seeking diabetes mellitus care in India: Evidence from a Nationally Representative Sample Survey. *Health Care Science, 2*(5), 291–305.

Nandi, S., Schneider, H., & Dixit, P. (2017). Hospital utilization and out of pocket expenditure in public and private sectors under the universal government health insurance scheme in Chhattisgarh State, India: Lessons for universal health coverage. *PLoS ONE, 12*(11), e0187904.

Nandi, S., & Schneider, H. (2020). When state-funded health insurance schemes fail to provide financial protection: An in-depth exploration of the experiences of patients from urban slums of Chhattisgarh. *India. Global Public Health, 15*(2), 220–235.

Oxfam India. (2021). Analysing regulation of private healthcare in India. Available: https://www.oxfamindia.org/knowledgehub/workingpaper/analysing-regulation-private-healthcare-india. Date Accessed: September 02, 2022.

Panda, B., & Rout, H. S. (2018). Extent of use of the government sponsored health insurance schemes: Evidence from rural Odisha. *Indian Journal of Economics and Development, 6*(10), 1–11.

Patel, V., Chatterji, S., Chisholm, D., et al. (2011). Chronic diseases and injuries in India. *The Lancet, 377*(9763), 413–428.

Philip, N. E., Kannan, S., & Sarma, S. P. (2016). Utilization of comprehensive health insurance scheme, Kerala: A comparative study of insured and uninsured below-poverty-line households. *Asia Pacific Journal of Public Health, 28*(1_suppl), 77S-85S.

Prinja, S., Bahuguna, P., Gupta, I., et al. (2019). Role of insurance in determining utilization of healthcare and financial risk protection in India. *PLoS ONE, 14*(2), e0211793. https://doi.org/10.1371/journal.pone.0211793

Prinja, S., Chauhan, A. S., Karan, A., Kaur, G., & Kumar, R. (2017). Impact of publicly financed health insurance schemes on healthcare utilization and financial risk protection in India: A systematic review. *PLoS ONE, 12*(2), e0170996.

Rahman, T., Gasbarro, D., & Alam, K. (2022). Financial risk protection from out-of-pocket health spending in low-and middle-income countries: A scoping review of the literature. *Health Research Policy and Systems, 20*(1), 1–23.

Rajasekhar, D., Berg, E., Ghatak, M., Manjula, R., & Roy, S. (2011). Implementing health insurance: The rollout of Rashtriya Swasthya Bima Yojana in Karnataka. *Economic and Political Weekly, 46*(20), 56–63.

Ranjan, A., Dixit, P., Mukhopadhyay, I., & Thiagarajan, S. (2018). Effectiveness of government strategies for financial protection against costs of hospitalization Care in India. *BMC Public Health, 18*(1), 1–12.

Rao, M., Katyal, A., Singh, P. V., Samarth, A., Bergkvist, S., Kancharla, M., ... Renton, A. (2014). Changes in addressing inequalities in access to hospital care in Andhra Pradesh and Maharashtra states of India: A difference-in-differences study using repeated cross-sectional surveys. *BMJ Open, 4*(6), e004471.

Raza, W., Van de Poel, E., & Panda, P. (2016). Analyses of enrolment, dropout and effectiveness of RSBY in northern rural India. *MPRA Paper*, 70081. Available: https://hsrii.org/wp-content/uploads/2016/07/MPRA_paper_70081.pdf

Rent, P., & Ghosh, S. (2015). Understanding the 'cash-less', nature of government-sponsored health insurance schemes: Evidence from Rajiv Gandhi Jeevandayee Aarogya Yojana in Mumbai. *SAGE Open, 5*(4), 1–10.

Reshmi, B., Unnikrishnan, B., Rajwar, E., Parsekar, S. S., Vijayamma, R., & Venkatesh, B. T. (2021). Impact of public-funded health insurances in India on health care utilisation and financial risk protection: A systematic review. *British Medical Journal Open, 11*(12), e050077.

Rout, S. K., Sahu, K. S., & Mahapatra, S. (2021). Utilization of health care services in public and private healthcare in India: Causes and determinants. *International Journal of Healthcare Management, 14*(2), 509–516.

Sabharwal, N. S., Mishra, V. K., Naik, A. K., Holmes, R., & Hagen-Zanker, J. (2014). How does social protection contribute to social inclusion in India, evidence from the National health insurance programme (RSBY) in Maharashtra and Uttar Pradesh. *Overseas development Institute*. Available: https://cdn.odi.org/media/documents/8938.pdf

Sahoo, A. K., & Madheswaran, S. (2014). Out-of-pocket (OOP) financial risk protection: The role of health insurance. *Institute for Social and Economic Change*. Available: https://103.154.2.116/~max233devisec/wp-content/uploads/2023/07/WP-330-Amit-Kumar-Sahoo_3.pdf

Sangar, S., Dutt, V., & Thakur, R. (2022). Burden of out-of-pocket health expenditure and its impoverishment impact in India: Evidence from National Sample Survey. *Journal of Asian Public Policy, 15*(1), 60–77. https://doi.org/10.1080/17516234.2019.1601065

Sangar, S., Dutt, V., & Thakur, R. (2019). Economic burden, impoverishment, and coping mechanisms associated with out-of-pocket health expenditure in India: A disaggregated analysis at the state level. *The International Journal of Health Planning and Management, 34*(1), e301–e313.

Selvaraj, S., & Karan, A. K. (2012). Why publicly-financed health insurance schemes are ineffective in providing financial risk protection. *Economic and Political Weekly, 47*(11), 60–68.

Selvaraj, S., Karan, A., Srivastava, S., Bhan, N., et al. (2022). India: Health system review. Health Systems in Transition, 11 (1), World Health Organization. Regional Office for South-East Asia. Available: https://apps.who.int/iris/handle/10665/352685

Sengupta, R., & Rooj, D. (2019). The effect of health insurance on hospitalization: Identification of adverse selection, moral hazard and the vulnerable population in the Indian healthcare market. *World Development, 122*, 110–129.

Sinha, R. K. (2018). Impact of publicly financed health insurance scheme (Rashtriya Swasthya bimA Yojana) from equity and efficiency perspectives. *Vikalpa, 43*(4), 191–206.

Smith, O., Dong, D., & Chhabra, S. (2019). PM-JAY and India's Aspirational Districts. National Health Authority; 2019 Sep p. 6. (PM-JAY Policy Briefs). Report No.: 3. Available: https://pmjay.gov.in/sites/default/files/2019-10/Policy%20Brief%203_PMJAY%20and%20India's%20Aspirational%20Districts%20(PRINT)_WB.pdf

Sood, N., Bendavid, E., Mukherji, A., Wagner, Z., Nagpal, S., & Mullen, P. (2014). Government health insurance for people below poverty line in India: Quasi-experimental evaluation of insurance and health outcomes. *Bmj, 349*. https://doi.org/10.1136/bmj.g5114

Sriram, S., & Khan, M. M. (2020). Effect of health insurance program for the poor on out-of-pocket inpatient care cost in India: Evidence from a nationally representative cross-sectional survey. *BMC Health Services Research, 20*(1), 1–21.

StataCorp, L. P. (2015). Stata statistical software: Release 14. Available: https://www.stata.com/news/14-1/

Taneja, P. K., & Taneja, S. (2016). Rashtriya Swasthya Bima Yojana (RSBY) for universal health coverage. *Asian Journal of Management Cases, 13*(2), 108–124.

Thakur, H. (2016). Study of awareness, enrollment, and utilization of Rashtriya Swasthya Bima Yojana (national health insurance scheme) in Maharashtra, India. *Frontiers in public health, 3*(282). Available: https://doi.org/10.3389/fpubh.2015.00282

WHO & The World Bank. (2017). Tracking universal health coverage: 2017 global monitoring report. Available: https://iris.who.

int/bitstream/handle/10665/259817/9789241513555-eng.pdf?sequence=1. Date Accessed: July 8, 2022.

Wagstaff, A., & van Doorslaer, E. (2003). Catastrophe and impoverishment in paying for health care: With applications to Vietnam 1993–1998. *Health Economics, 12*, 921–933. https://doi.org/10.1002/hec.776

Wagstaff, A., Eozenou, P., & Smitz, M. (2020). Out-of-pocket expenditures on health: A global stocktake. *The World Bank Research Observer, 35*(2), 123–157.

Watson, J., Yazbeck, A. S., & Hartel, L. (2021). Making health insurance pro-poor: Lessons from 20 developing countries. *Health Systems & Reform, 7*(2), e1917092.

World Health Organization (WHO). (2014). WHO Global Health Expenditure Atlas. Geneva: Department of Health Systems Financing Health Systems Services, WHO. Available: https://cdn.who.int/media/docs/default-source/health-financing/atlas2014.pdf . Date Accessed: 7 July, 2023.

World Health Organization (WHO). (2020a). Global Health Expenditure Database. Available: https://apps.who.int/nha/database/country_profile/Index/en. Date Accessed: Jan 27, 2023.

World Health Organization and The World Bank. (2021). Global monitoring report on financial protection in health 2021. License: CC BY-NC-SA 3.0 IGO. Available: https://www.who.int/publications/i/item/9789240040953. Date Accessed: January 01, 2023.

World Health Organization. (2020b). Primary health care on the road to universal health coverage: 2019 Global Monitoring Report. Available: https://www.who.int/publications/i/item/9789240029040. Date Accessed: September 10, 2022.

Yadav, S., & Arokiasamy, P. (2014). Understanding epidemiological transition in India. *Global Health Action, 7*(1), 23248. https://doi.org/10.3402/gha.v7.23248

Yadav, J., John, D., Menon, G. R., & Singh, R. (2024). Gender inequality in out-of-pocket expenditure on healthcare among persons with disabilities in India. *Journal of Health Management, 26*(2), 238–255.

Zhao, Q. Y., Luo, J. C., Su, Y., Zhang, Y. J., Tu, G. W., & Luo, Z. (2021). Propensity score matching with R: Conventional methods and new features. *Annals of translational medicine, 9*(9), 812.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Journal Pre-proof



Improved optical characteristics in $BaBi_2Nb_2O_9$ ferroelectric ceramic infused with transition metal ion (W^{6+}) and rare earth ions (Er^{3+}/Yb^{3+})

Ankita Banwal, Bhavya Kumar, Manoj Verma, Ankur Shandilya, Bharti Singh, Renuka Bokolia

PII: S0022-2313(24)00373-9

DOI: https://doi.org/10.1016/j.jlumin.2024.120809

Reference: LUMIN 120809

To appear in: *Journal of Luminescence*

Received Date: 20 May 2024

Revised Date: 27 June 2024

Accepted Date: 19 July 2024

Please cite this article as: A. Banwal, B. Kumar, M. Verma, A. Shandilya, B. Singh, R. Bokolia, Improved optical characteristics in $BaBi_2Nb_2O_9$ ferroelectric ceramic infused with transition metal ion (W^{6+}) and rare earth ions (Er^{3+}/Yb^{3+}), *Journal of Luminescence*, https://doi.org/10.1016/j.jlumin.2024.120809.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2024 Published by Elsevier B.V.

Journal Pre-proof

Improved optical characteristics in $BaBi_2Nb_2O_9$ ferroelectric ceramic infused with transition metal ion (W^{6+}) and rare earth ions (Er^{3+}/Yb^{3+})

Ankita Banwal[1], Bhavya Kumar[1], Manoj Verma[2], Ankur Shandilya[2], Bharti Singh[1], Renuka Bokolia[1*]
[1]*FMRL, Department of Applied Physics, Delhi Technological University, India*
[2]*Department of Physics, Hindu College, University of Delhi, India*

Abstract: In this paper, a detailed study of the Er/Yb/W tri-doped $BaBi_{2-x-y}Nb_{2-z}Er_xYb_yW_zO_9$ (BBN) system and a comparison of undoped, Er-doped, Er/Yb co-doped and Er/Yb/W tri-doped BBN systems are presented. All the BBN ceramics are prepared using the solid-state approach. The XRD spectra demonstrated that each BBN composition has an orthorhombic geometry. The UCL spectra are traced at 980 nm and exhibit two distinctive green bands and one red band at 535 nm, 556 nm, and 673 nm, respectively. The UCL intensity improves with W^{6+} concentration and reaches a maximum at (z = 0.02); after this concentration, the quenching effect occurs. The pump power variation on UCL spectra confirmed the involvement of two photons. The time-resolved fluorescence analysis was done for green and red emissions, and it was observed that the time decay increased for higher dopant concentrations, suggesting a non-radiative transition at higher dopant concentrations. The FIR technique measured the absolute sensitivity and had a value of 0.91% K^{-1} at 573 K. These improvements suggest that the prepared material may help develop advanced electronic equipment for multifunctional applications.

Keywords: Upconversion; time decay; sensitivity; BBN; ferroelectric; ceramic.

1. Introduction

Ferroelectric materials have gained prominence due to their multifunctional applications in optics, electrical engineering, bioimaging, sensing, energy storage, and many more [1-5]. They have high chemical and thermal stability, which is required to modify the host material by inserting foreign ions. Furthermore, the phonon energy of these materials is low, which is sufficient for the upconversion mechanism. The advances in electrical and optical applications have shown tremendous possibilities. Researchers have been inspired to explore multifunctional and hybrid materials due to the need for cost-effective and environmentally friendly materials capable of fulfilling several purposes [6-8]. The increasing recognition of layered bismuth structure ferroelectric (BLSF) materials may be attributed to their advantageous chemical and thermal durability, exceptional radiation tolerance, reduced phonon energy, and enhanced sensing capacity over a wide temperature range. BLSF provides a compelling alternative to lead-based ceramics due to their non-toxic nature and ecologically friendly characteristics, making them a preferable replacement choice. The BLSF comprise a distinct category of ferroelectric materials and are denoted by the formula $(Bi_2O_2)^{2+}$ $(A_{n-1}B_nO_{3n+1})^2$. The quasi-perovskite layer, $A_{n-1}B_nO_{3n+1}$, is enveloped by two layers of bismuth oxide $(Bi_2O_2)^{2+}$. Here, n is the number of octahedral layers that comprise the perovskite unit, A denotes mono, di, and tri valence ions, and B symbolizes higher valence ions (tetra, penta, hexa) [9-12].

The literature analysis revealed that incorporating rare earth (RE) ions into BLSF materials improved their electrical, optical, and sensing characteristics. In recent years, temperature sensing has attracted considerable interest due to breakthroughs in business and technology. Non-contact temperature sensors are needed in power plant electrical transformers in oil and coal factories, where traditional temperature monitoring equipment is prohibited. Developing high-temperature sensing materials with a substantial degree of sensitivity poses an immense challenge for contemporary researchers studying RE-doped materials. Luo et al., for instance, investigated the efficacy of Er^{3+}/Yb^{3+} co-doped $Ba_2NaNb_5O_{15}$ glass ceramics regarding UC luminescence, thermal sensing, and energy storage [13]. Additionally, a prior study employed co-doped tungsten tellurite glass doped with Er^{3+} and Yb^{3+} to fabricate a temperature sensor [14]. The sensor exhibited a respectable green emission luminescence and sensitivities of approximately 28.72×10^{-4} K^{-1} at 690 K. The luminescence intensity ratio of La_2TiO_7 ceramics doped with Yb^{3+}/Er^{3+} was investigated by Swami et al. [15]. At 498 K, the authors obtained a sensitivity of 0.29% K^{-1}. Furthermore, in $Bi_{0.5}Na_{0.5}TiO_3$: Yb^{3+}/Er^{3+}, approximate sensing of 0.49% K^{-1} at 363 K within an extensive temperature spectrum has been reported [16]. From the literature survey, it was found that to refine the luminous characteristics along with the electrical properties of lanthanide-doped materials, it is possible to alter the surroundings of the emission centers. The inclusion of some non-lanthanide elements has shown encouraging outcomes in effectively altering the nearby crystal field around the rare-earth ions. Several non-lanthanide elements, including Li^+, Mg^{2+}, Zn^{2+}, Mo^{3+}, Bi^{3+}, and W^{6+}, have been reported to have the ability to enhance the optical as well as electrical properties. For instance, the electrical and structural characteristics of Bi_3TiNbO_9 ferroelectric ceramic are improved by the incorporation of W^{6+} ions [17]. Hou.

* Corresponding Author
Email: renukabokolia@dtu.ac.in

et. al. reported the remanent polarization and d_{33} coefficient were increased in W^{6+} doped $CaBi_2Nb_2O_9$ [18]. In $Ho^{3+}/Yb^{3+}/Zn^{2+}$ tri-doped Y_2O_3 phosphor, the UCL emission is enhanced 11 times [19]. In $Bi_4Ti_3O_{12}$ ferroelectric ceramic, the upconversion luminescence and optical temperature sensitivity were improved by the introduction of W^{6+} ions along with Er^{3+} and Yb^{3+} ions [20].

The objective of this paper is to conduct a comprehensive investigation of the impact of singly doped Er^{3+} ions, co-doped Er^{3+}/Yb^{3+} ions, and tri-doped $Er^{3+}/Yb^{3+}/W^{6+}$ ions on the $BaBi_2Nb_2O_9$ ferroelectric ceramic. Enhanced green upconversion emission was observed in the ferroelectric ceramic $BaBi_2Nb_2O_9$ doped with Er^{3+} in our previous investigation. The highest upconversion luminescence (UCL) was noted at an Er^{3+} content of x = 0.04 [21]. Then, in another study, the Er^{3+} doped $BaBi_2Nb_2O_9$ ferroelectric ceramic is optimized and co-doped with Yb^{3+} ions. The highest green emission was observed at a Yb^{3+} (y = 0.10) concentration [22]. The impact of Er^{3+} and Yb^{3+} ions on electrical properties was also examined in a separate study [23]. These discoveries served as an incentive for us to further investigate the $BaBi_2Nb_2O_9$ system in order to enhance its properties by tri-doping it with $Er^{3+}/Yb^{3+}/W^{6+}$ for multifunctional applications. Therefore, in this paper, we conducted a systematic replacement of a B-site ion (Nb^{5+}) with a transition metal ion (W^{6+}) to investigate the optical and sensing properties of $BaBi_2Nb_2O_9$.

2. Synthesis and characterization description

The ceramic $BaBi_{2-x-y}Nb_{2-z}Er_xYb_yW_zO_9$ (BBN) is manufactured by a simple solid-state method. The Er^{3+} (x = 0.04) and Yb^{3+} (y = 0.10) were fixed at constant concentrations, and W^{6+} is varied during the process (z = 0.01 to 0.04). The selection of Er^{3+} (x = 0.04) and Yb^{3+} (y = 0.10) was based on the fact that this combination of dopant concentrations exhibited the highest green emission intensity peak in our previous work [21, 22]. The precursors used were $BaCO_3$, Bi_2O_3, Nb_2O_5, Er_2O_3, and Yb_2O_3, WO_3 having the highest purity (99.99%) from Sigma Aldrich. These initial substances were measured in stoichiometric proportions and homogenized using a mortar and pestle for 5 hours, using ethanol as the solvent. The automated high-temperature furnace is used for the calcination of the sample at 950°C with a hold time of 3 hours. Then, a binder PVA is mixed with calcined powder to prepare the pellets. The pellets were fabricated using a hydraulic press at a pressure of 5 MPa, having a diameter of 10 mm and a thickness of approximately 1 mm. These pellets were again put into the same furnace for sintering at 1050°C for 3 hours for the sample to harden. The schematic description of the preparation method via solid state route is given in Fig. 1.

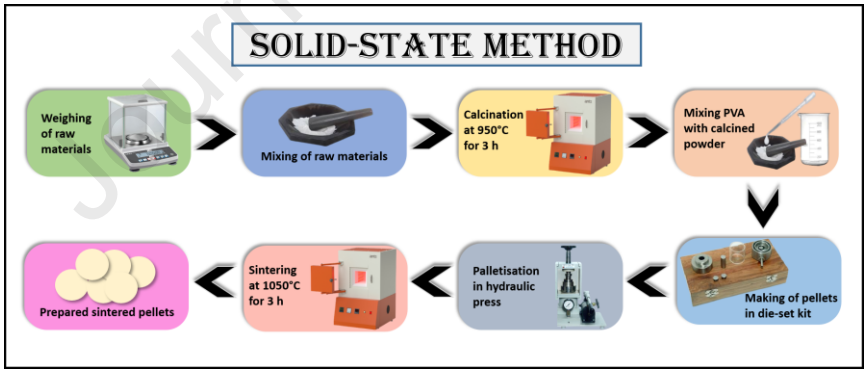


Fig. 1. Steps involved in solid-state method.

Table 1: Abbreviation used for various compositions of $BaBi_{2-x-y}Nb_{2-z}Er_xYb_yW_zO_9$ ceramic.

Compositions	Sample Code	Er^{3+} (x)	Yb^{3+} (y)	W^{6+} (z)
$BaBi_{2-x-y}Nb_{2-z}Er_xYb_yW_zO_9$	BBNE0Y0W0	0.00	0.00	0.00
$BaBi_{1.96}Nb_{2-z}Er_{0.04}Yb_{0.00}W_{0.00}O_9$	BBNE4Y0W0	0.04	0.00	0.00
$BaBi_{1.86}Nb_{2-z}Er_{0.04}Yb_{0.10}W_{0.00}O_9$	BBNE4Y10W0	0.04	0.10	0.00
$BaBi_{1.86}Nb_{1.99}Er_{0.04}Yb_{0.10}W_{0.01}O_9$	BBNE4Y10W1	0.04	0.10	0.01
$BaBi_{1.86}Nb_{1.98}Er_{0.04}Yb_{0.10}W_{0.02}O_9$	BBNE4Y10W2	0.04	0.10	0.02
$BaBi_{1.86}Nb_{1.97}Er_{0.04}Yb_{0.10}W_{0.03}O_9$	BBNE4Y10W3	0.04	0.10	0.03
$BaBi_{1.86}Nb_{1.96}Er_{0.04}Yb_{0.10}W_{0.04}O_9$	BBNE4Y10W4	0.04	0.10	0.04

The investigation of the phase of the material was conducted through the analysis of XRD spectra utilizing Cu-Kα radiations by the use of a Bruker high-quality X-ray diffractometer. The refinement and lattice parameters are calculated by TOPAZ software. The HORIBA quanta master instrument was used to facilitate UCL spectra using laser source light at 980 nm. For the lifetime decay measurement, a pulsed laser was operated in modulation mode at its maximum power capability of 2W/cm^2 for 30 minutes. The synthesized

material was placed on the heaters to analyze the temperature-dependent UCL emission spectra while the temperature was measured with a thermocouple. A manually constructed heater equipped with a thermocouple, multimeter, and variable-voltage control was utilized to conduct the sensing, resolution, and stability analyses. The sample coding for all the prepared BBN ceramics is recorded in Table 1.

3. Results and discussion

3.1. XRD analysis

The spectra of XRD for undoped, Er^{3+} doped, Er^{3+}/Yb^{3+} co-doped, and $Er^{3+}/Yb^{3+}/W^{6+}$ tri-doped BBN ceramic is represented in Fig. 2. The X-ray diffraction (XRD) peaks exhibit a notable resemblance to the JCPDF file number 00-012-0403, which is associated with an orthorhombic crystal structure and is classified within the Fmmm phase group. The absence of secondary peaks in the XRD patterns provides evidence for an extensive incorporation of the dopant within the host lattice. The prominent XRD peak corresponds to the (115) plane, attributed to the BLSFs having n = 2. From the previous findings, it can be confirmed that the (112n+1) plane is associated with an intense XRD peak in the BLSF materials of the Aurivillius family. The density of the prepared sintered pellet is evaluated by the Archimedes method (Table 2). The measured density is comparable with the reported density in the JCPDS file with 95% accuracy. The A-site Bi^{3+} ions are replaced by Er^{3+} and Yb^{3+} ions at an optimized concentration, whereas W^{6+} ions systematically replace the B-site ion Nb^{5+}. The explanation for this replacement is based on the fact that they have equal charge valency and nearly comparable radii. The resulting change in lattice parameters is shown in Table 2, indicating a minor shift in the parameters, which implies the maintenance of electroneutrality inside the lattice, which prevents lattice deformation. Also, the increment in lattice cell volume can be explained by the lone pair character of Bi^{3+} ions. In the $BaBi_2Nb_2O_9$ structure, this phenomenon might be attributed to the restricted nature of the $6s^2$ lone pair of Bi^{3+} inside the host. This restricted lone pair in Bi^{3+} induces a little deformation in the surrounding lattice structure, resulting in an expansion of the unit cell volume [6, 15, 24].

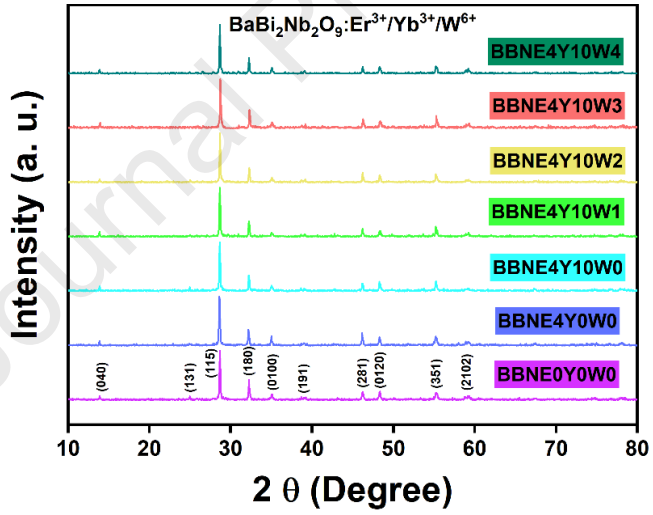


Fig. 2. XRD spectra of all the doped and undoped BBN ceramics compositions.

Table 2: Refinement parameters, lattice parameters, and experimental density of all the compositions of BBN.

Parameters	BBNE0Y0W0	BBNE4Y0W0	BBNE4Y10W0	BBNE4Y10W1	BBNE4Y10W2	BBNE4Y10W3	BBNE4Y10W4
R_p	12.84	13.59	11.76	10.56	10.93	10.37	11.02
R_{wp}	11.00	12.18	12.98	9.22	9.38	8.74	9.57
R_{exp}	10.76	11.57	10.22	8.52	8.66	8.32	8.82
R Braggs	13.606	9.391	9.651	9.971	11.566	12.757	8.886
Gof (χ)	1.02	1.05	1.06	1.08	1.08	1.05	1.08
a (Å)	5.5394	5.5526	5.5488	5.5675	5.5640	5.5631	5.5627
b (Å)	25.5751	25.609	25.620	25.6679	25.6734	25.6771	25.6625
c (Å)	5.5571	5.5623	5.5612	5.5637	5.56301	5.5658	5.5577
V ($Å^3$)	787.294	790.982	790.581	795.095	794.672	795.063	793.394
Orthorhombicity	1.0031	1.0017	1.0022	1.0006	1.0001	1.0004	1.0008
Density (g/cm^3)	7.0091	6.8724	6.7600	6.1919	6.2246	6.3253	6.6633

3.2 Light upconversion luminescence spectra analysis

The upconversion luminescence (UCL) spectra of W^{6+} doped BBN compositions for different W^{6+} concentrations (z = 0.01, 0.02, 0.03, and 0.04) are shown in Fig. 3(a). The spectra were acquired using an irradiation of 980 nm excitation wavelength with a source density of 3.25 W/cm^2. At 535 nm and 546 nm, two distinct green emission bands were detected due to the transitions from $^4H_{11/2}$ to $^4I_{15/2}$ and $^4S_{3/2}$ to $^4I_{15/2}$,

respectively. The red band at 673 nm through the transition from $^4F_{9/2}$ to $^4I_{15/2}$ can be visible. The crystal field causes the $^4F_{9/2}$ energy level to be split into several sublevels via upconversion luminescence. Fig. 3(b) demonstrates that the emission intensity reaches its maximum value at a concentration of z = 0.02. Subsequently, the emission intensity decreases due to the occurrence of the concentration quenching effect. In this effect, the proximity of the dopant ions becomes sufficiently close, leading to their involvement in non-radiative transitions, hence causing a deterioration in the photoluminescence. The calculation of the interionic distance between dopant ions can be accomplished by the Blasse equation (1) [25]:

$$R_c \approx 2\left[\frac{3V}{4\pi X_c Z}\right]^{\frac{1}{3}} \quad (1)$$

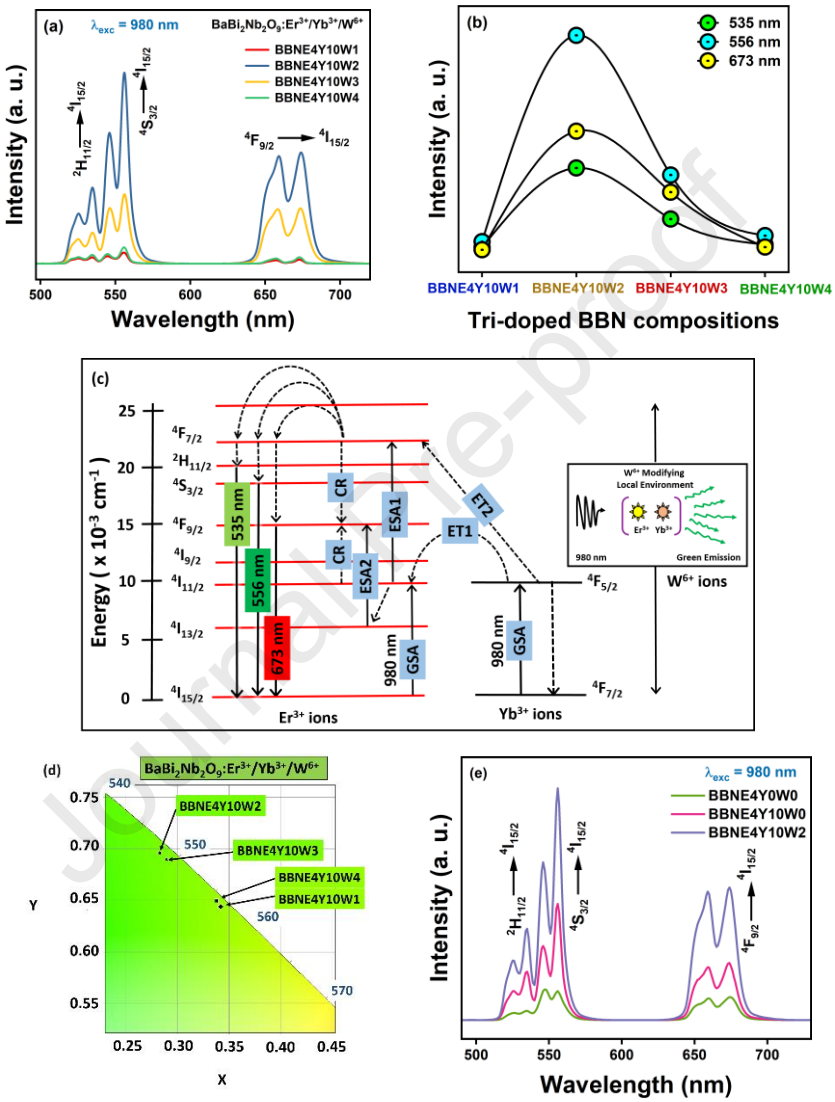


Fig. 3. (a) UCL spectra of all the compositions of tri-doped BBN ceramics (b) Variation of UCL intensity in the tri-doped system (c) Schematic depiction of energy levels of RE^{3+} ions (d) CIE (x, y) plot of tri-doped BBN ceramics (e) Comparison of UCL intensity of BBNE4Y0W0, BBNE4Y10W0, and BBNE4Y10W2 ceramic.

Here, V represents the volume of the unit cell, Z denotes the number of host cations, and X_c (x+y+z) refers to the optimal concentration at which quenching takes place. By substituting the given values into the equation mentioned above, the value of R_c may be determined as 13.34 Å. Fig. 3(c) depicts the energy level diagram illustrating the potential pathways for the up-conversion luminescence (UCL). The Yb^{3+} ions are subjected to illumination from a 980 nm laser, which results in energy absorption by ground state absorption (GSA). This energy is transferred to a higher energy level at $^4I_{11/2}$ through energy transfer 1 (ET1) and $^4F_{7/2}$ through energy transfer 2 (ET2). Photons may potentially be absorbed by Er^{3+} ions via GSA. Therefore, the level $^4I_{11/2}$ is produced by the ESA (excitation by stimulated emission) process to transition to a higher energy level, namely the $^4F_{7/2}$ level. Consequently, the level denoted as $^4F_{7/2}$ experiences an excessive population due to two distinct processes: energy transfer (ET) and excited state absorption (ESA). The ions undergo a non-radiative process known as multiphonon relaxation (MPR), resulting in their transition to energy levels $^2I_{11/2}$, $^4S_{3/2}$, and $^4F_{9/2}$.

Ions emit a single red band at 672 nm and two green bands at 535 and 556 nm due to their transition to the lower energy level $^4F_{15/2}$. The augmented red emission seen with a change in Yb^{3+} concentration may be attributed to an additional decay path facilitated via cross-relaxation channels CR1 and CR2. The CIE chromaticity coordinates (x, y) for different tri-doped BBN compositions were also calculated from the emission spectra. It has been observed that at z = 0.01 and 0.04, the spectra exhibit in the brighter green region, whereas for z = 0.02 and 0.03, they manifest in a deeper green region, as depicted in Fig. 3(d). Furthermore, the enhancement in UCL emission intensity by the incorporation of W^{6+} ions in the host lattice can be seen in Fig. 3(e). The local electric field surrounding the Er^{3+} ions changes, increasing the chance of electronic-dipole (ED) transitions, which causes this intensity spike. Consistent with earlier observations, the results show that changes in the crystal field influence the UC emissions of Er^{3+} ions linked to the intra-4f electronic dipole transitions [26].

3.3. Impact of changing pump power on the UCL emission spectra

To get a deeper understanding of the mechanisms behind upconversion luminescence, we investigated the transitions involving the levels $^2H_{11/2}$, $^4S_{3/2}$, and $^4F_{9/2}$ to level $^4I_{15/2}$ in the BBNE4Y10W2 composition system. The distinct pump powers range from 89 mW to 389 mW with an excitation source density of 5.79 W/cm^2 to 25.34 W/cm^2, as depicted in Fig. 4(a). It has been discovered that as the pump power increases, the UCL intensity also increases. Eq. (2) illustrates the correlation between pump power and UCL emission intensity [27, 28].

$$I_{UP} \propto P^n \tag{2}$$

Here, P represents the pump power, n signifies the low-energy photons necessary for the UCL, and I_{UP} denotes the UCL intensity. The curve of ln(intensity) against the ln(Power) is given in Fig. 4(b). The value of n (no. of photons involved in UC) can be obtained from the graph's slope. The n equals 1.33, 1.42, and 1.20 for emission bands at 535 nm, 556 nm, and 673 nm, respectively. The fact that n is greater than 1 means that more than one photon is involved in UCL. The disparity from n = 2 might be seen due to the impact of distinct UC mechanisms [26].

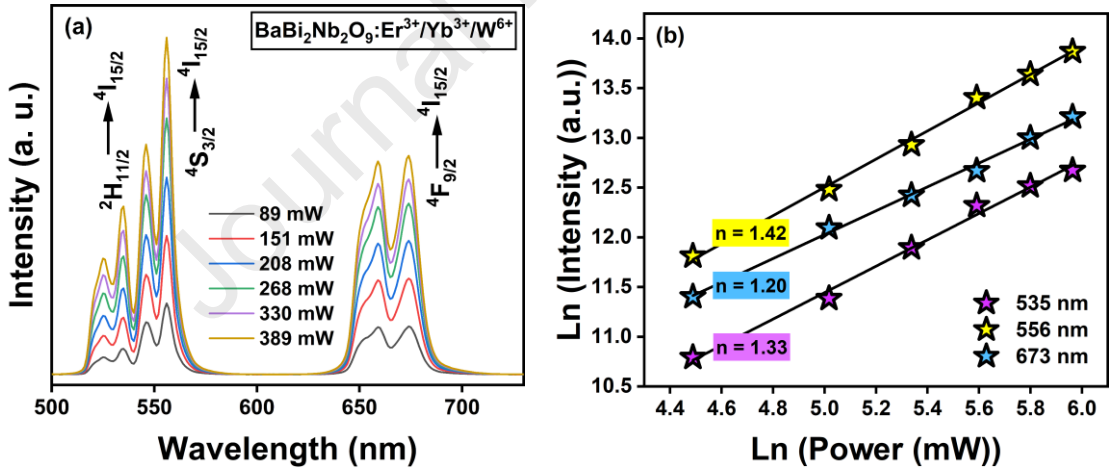


Fig. 4. (a) Dependency of varying pump power on UCL spectra (b) Ln(Intensity) versus Ln(Power).

3.4. Time-resolved photoluminescence analysis of green and red emission bands

In time-dependent dynamics of luminescence, the emission intensity gradually increases until it reaches its highest point as the excitation is prolonged. After the excitation source is removed, the emission intensity gradually decreases to zero, resulting in a comprehensive time-resolved curve. In general, the time required for an emission to decay to the 1/e value of the initial intensity is defined as the luminescence lifetime. In this study, the laser source in the modulation mode under the excitation of 980 nm wavelength is used for the time-resolved photoluminescence spectra of the green band and red band through transition $^4S_{3/2}$ to $^4I_{15/2}$ and $^4F_{9/2}$ to $^4I_{15/2}$ energy levels, respectively. The time decay profile of the green band at 556 nm for the Er^{3+} doped, Er^{3+}/Yb^{3+} co-doped, and $Er^{3+}/Yb^{3+}/W^{6+}$ tri-doped BBN system is given in Fig. 5(a-f). The curve is fitted using decay exponential fit given by Eq. (3):

$$I(t) = I_0 + \sum_i A_i \exp(-\frac{t}{\tau_i}) \ (i = 1, 2, 3) \tag{3}$$

where A_i (i = 1, 2, 3) is the fitting coefficient, τ_i (i = 1, 2, 3) is the decay time of luminescence, and I_0 is the intensity at the initial time. The mean lifetime is calculated using Eq. (4):

$$< \tau > = \sum_i \frac{A_i \tau_i^2}{A_i \tau_i} \text{ (i = 1, 2, 3)} \quad (4)$$

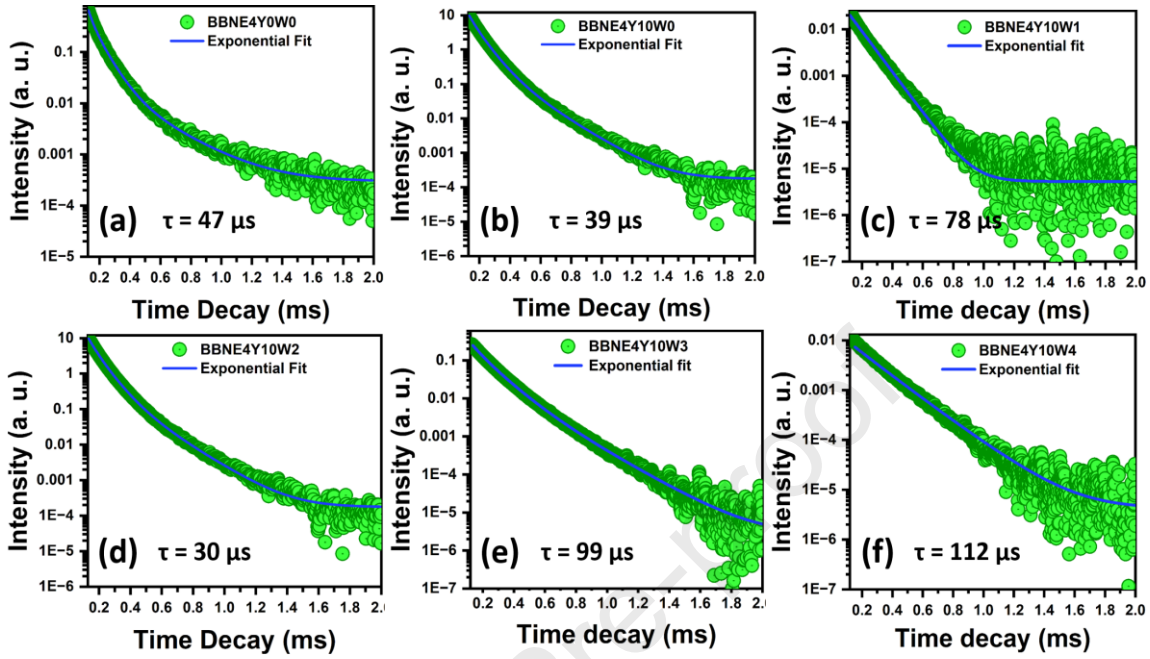


Fig. 5. Time decay curve of (a) BBNE4Y0W0 (b) BBNE4Y10W0 (c) BBNE4Y10W1 (d) BBNE4Y10W2 (e) BBNE4Y10W3 (f) BBNE4Y10W4 ceramics of green emission band at 556 nm.

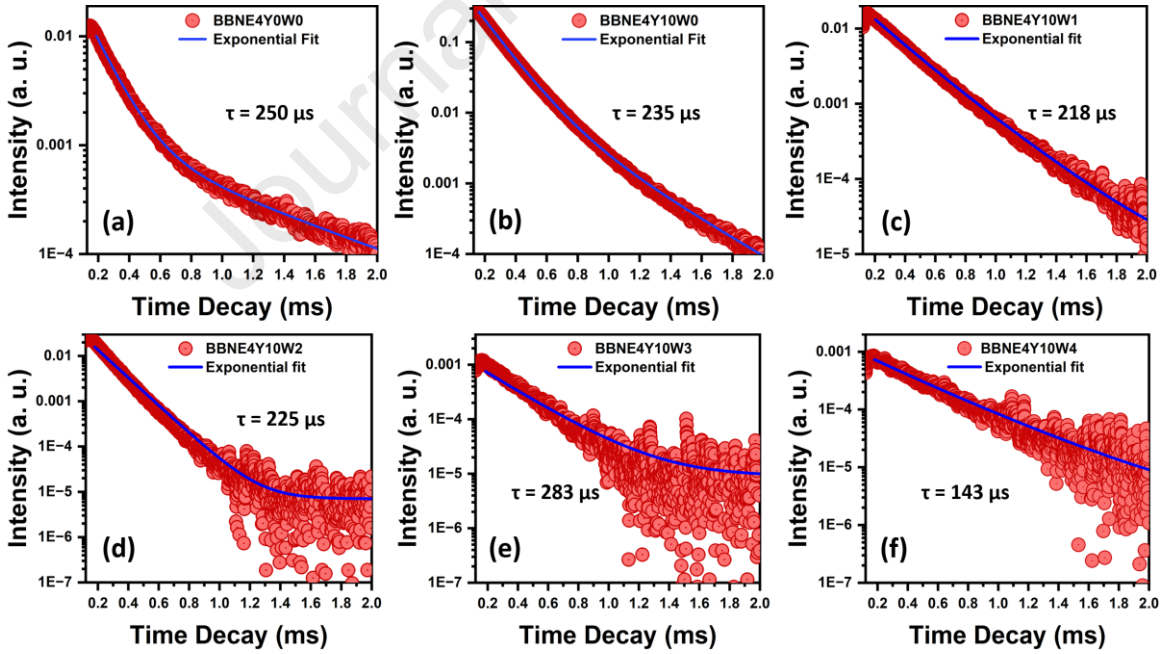


Fig. 6. Time decay curve of (a) BBNE4Y0W0 (b) BBNE4Y10W0 (c) BBNE4Y10W1 (d) BBNE4Y10W2 (e) BBNE4Y10W3 (f) BBNE4Y10W4 ceramics of red emission band at 673 nm.

The average lifetime for optimized (z = 0.02) BBNE4Y10W2 composition is 30 µs, which supports the high emission UCL intensity. After this, the average lifetime increases, possibly due to the non-radiative transitions at higher dopant concentrations. It has also been observed that the average lifetime of the green emission band at 556 nm is decreasing from single-doped BBN to tri-doped BBN, which prompts that the tri-doped system has a larger transition probability of the emitting level than the single-doped system [29]. Further, the time-resolved spectroscopy for the red emission band through a transition of ions from $^4F_{9/2}$ to $^4I_{15/2}$ is shown in Fig. 6(a-f). The average decay time for red emission is increasing with increasing W^{6+} concentration. The average

lifetime of the red emission band is also reduced in the co-doped and tri-doped systems due to the effective radiative transition probabilities in these BBN compositions. Also, the presence of Yb^{3+} ions prevents the clustering of Er^{3+} ions, thereby increasing the optical properties [30].

3.5. Contactless temperature sensitivity using FIR technique

The well-known method for contactless temperature sensitivity is the fluorescence intensity ratio (FIR) technique. This method is compatible with checking the capabilities of materials to be used as optical temperature sensors. We calculated the sensitivity of optimized W^{6+} content (z = 0.02) for the applicability of ceramic to be used as potential sensor material. The temperature-dependent UCL spectra are plotted in Fig. 7(a). In this method, the ratio of the integral intensity of green emission bands via transitions $^4H_{11/2}$ to $^4I_{15/2}$ and $^4S_{3/2}$ to $^4I_{15/2}$ is depicted in Fig. 7(b) and is calculated by Eq. (5) [26]:

$$FIR = \frac{I_H}{I_S} = C\exp\left(-\frac{\Delta E}{K_B T}\right) \quad (5)$$

where I_H and I_S are the symbolic representation for intensities via transitions $^4H_{11/2}$ to $^4I_{15/2}$ and $^4S_{3/2}$ to $^4I_{15/2}$, respectively; ΔE is the difference between thermally coupled energy levels, C is the coefficient of exponential, K_B is the Boltzmann constant, and T is the temperature at which spectra is traced. Eq. (5) can be re-written in linear form to ΔE and C and is given in Eq. (6) [26]:

$$Ln(FIR) = -\left(\frac{\Delta E}{K_B}\right)\left(\frac{1}{T}\right) + LnC \quad (6)$$

The Ln(FIR) slope against the temperature inverse gives us the values of $\Delta E/K_B$ and LnC to be 1158.53 and 2.97, respectively, provided in Fig. 7(c). These parameters determine the optical temperature sensitivity of the produced ceramic BBNE4Y10W2. Eq. (7) may determine a material's absolute sensitivity by plotting its changing FIR value against temperature. [26]:

$$S_a = \frac{d(FIR)}{dT} = \frac{d\left(C\exp\left(-\frac{\Delta E}{K_B T}\right)\right)}{dT} = C\exp\left(-\frac{\Delta E}{K_B T}\right)\left(\frac{\Delta E}{K_B T^2}\right) \quad (7)$$

To acquire the relative sensitivity (S_r) of the material, the absolute sensitivity (S_a) is divided by FIR as employed in Eq. (8) [26]:

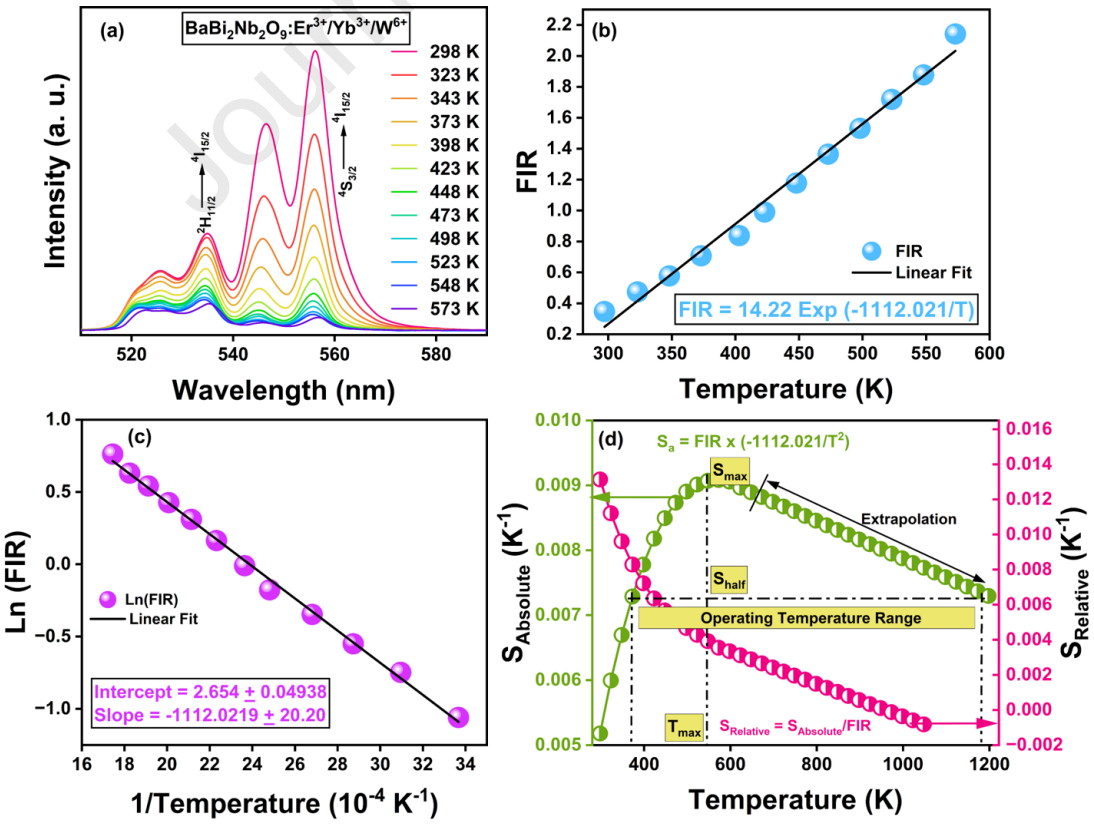


Fig. 7. (a) Emission pattern of BBNE4Y10W2 sample with varying temperature from 303-573 K (b) FIR versus temperature (c) Ln(FIR) against the inverse of temperature (d) Variation of S_a and S_r with temperature.

$$S_r = \frac{1}{FIR} * S_a = \left(\frac{\Delta E}{K_B T^2}\right) \quad (8)$$

The curve of S_r and S_a is shown in Fig. 7(d). The maximum S_a obtained at 573 K is 0.91% K^{-1}, whereas S_r is 1.32% K^{-1} at 297 K. The operational thermal range for sensors is represented by the shaded area in Fig. 9(d); S_a values are extrapolated up to S_{half} (S_a is more than half of the highest value) using Eq. 9. When the experimentally observed ΔE_{ex} and the theoretically calculated ΔE_{th} are nearly same, the material sensitivity is considered to be suitable. The ΔE_{ex} is calculated from the slope of Fig. 7(c), and the deconvoluted emission spectra give ΔE_{th} by the Viogt fit, as illustrated in Fig. 8. The inaccuracy is 2.8% calculated from Eq. (9) [24]:

$$\delta = \frac{|\Delta E_{th} - \Delta E_{ex}|}{\Delta E_{th}} * 100\% \quad (9)$$

The repeatability (R) and resolution (δT) of a sensing material are critical in determining its effectiveness. These parameters can be presented by Eqs. (10) and (11) [31]:

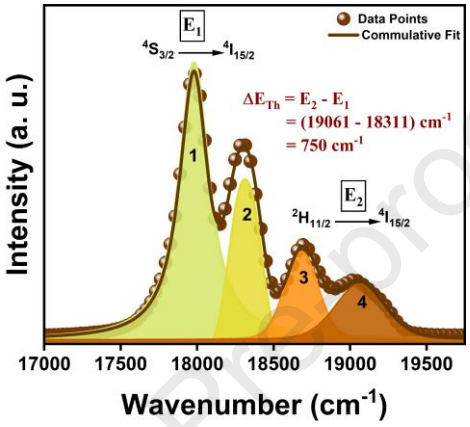


Fig. 8. Deconvoluted curve of BBNE4Y10W2 sample using Voigt fit for ΔE_{th} calculation.

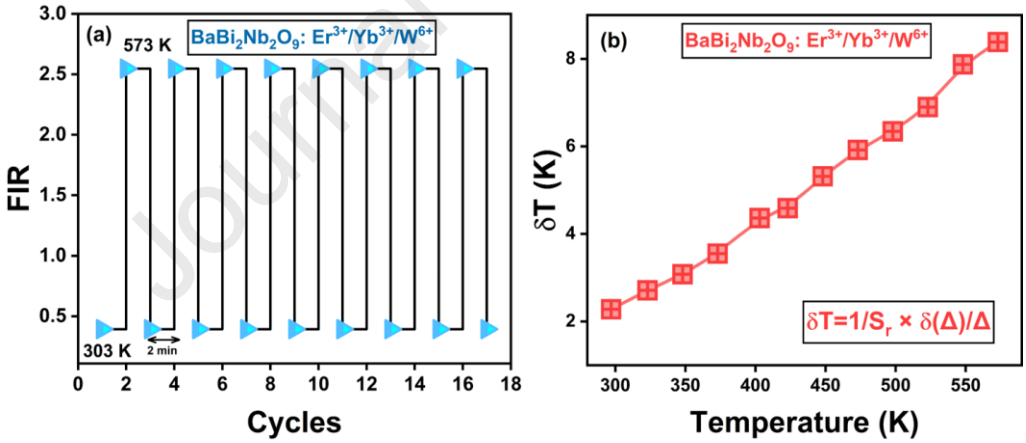


Fig. 9. (a) Repeatability test of BBNE4Y10W2 for 17 cycles with stabilization time of 2 minutes (b) Resolution of temperature from 303 K to 573 K.

$$R = 1 - \frac{\max(\Delta_{Av} - \Delta_{Sp})}{\Delta_{Av}} \quad (10)$$

$$\delta T = \frac{1}{S_r} \times \frac{\delta_\Delta}{\Delta} \quad (11)$$

where Δ_{Av} and Δ_{Sp} are FIR's mean and specific values in the temperature range 303-573 K for 17 cycles, and δ_Δ/Δ denotes the device's accuracy. Fig. 9(a) shows negligible change in the FIR values across the entire temperature range. The temperature resolution has improved from 2.67 K to 8.78 K, as depicted in Fig. 9(b). Table 3 compares the sensitivity of the Er^{3+} doped, Er^{3+}/Yb^{3+} doped, and $Er^{3+}/Yb^{3+}/W^{6+}$ doped BBN system with other RE-doped ferroelectric systems. Hence, it can be deduced that the prepared ceramic is a proficient material for sensing and has the potential to be extensively studied for its suitability in optical temperature sensor applications.

Table 3: Comparison of the sensitivity of various doped ferroelectric ceramics.

Host material	ΔE (cm^{-1})	Temperature Range (K)	T_a (K)	S_a (K^{-1})	S_r (K^{-1})	T_r (K)	References
$BaTiO_3$: Er/Yb	600	125-430	410	0.19%	-	-	[2]
$CaBi_2Ta_2O_9$: Er/Yb	1181	153-603	590	0.39%	-	-	[3]
$Na_{0.5}Er_{0.5}Bi_4Ti_4O_{15}$	500.1	175-500	380	0.17%	-	-	[11]
$Ba_2NaNb_5O_{15}$: Er/Yb	735.5	298-573	373	0.68%	0.90%	298	[13]
$La_2Ti_2O_7$: Er/Yb	683	93-773	493	0.29%	0.41 %	300	[15]
BNT-SBT: Er/Yb	-	303-573	363	0.49%	-	-	[16]
$BaBi_2Nb_2O_9$: Er	661	303-623	483	0.99%	0.40%	303	[21]
$BaBi_2Nb_2O_9$: Er/Yb (y = 0.06)	705	303-573	523	0.69%	1.10%	303	[22]
$BaBi_2Nb_2O_9$: Er/Yb (y = 0.10)	645	303-573	463	0.58%	1.01%	303	[22]
$Na_{0.5}Bi_{2.5}Ta_2O_9$: Er/Yb	-	153-483	483	0.78%	5.99%	153	[28]
$Na_{0.5}Bi_{0.5}TiO_3$: Er	706.68	173-553	493	0.35%	-	-	[31]
$BaBi_2Nb_2O_9$: Er/Yb/W	773	303-573	573	0.91%	1.32%	303	**This work**

4. Conclusion

The solid-state method is used to produce the ferroelectric ceramic $BaBi_{2-x-y}Nb_{2-z}Er_xYb_yW_zO_9$ (BBN), which is doped with the transition metal ion, W^{6+} at a constant concentration of Er^{3+} and Yb^{3+}. The X-ray diffraction (XRD) spectra revealed the presence of a perovskite phase in each ceramic sample, which exhibited an orthorhombic crystal structure. The UCL spectrum displayed two distinct green bands and one red band. The UCL intensity exhibits enhancement with an increasing concentration of W^{6+} ions, peaking at (z = 0.02). Beyond this concentration, a quenching effect occurs. The presence of two photons was verified by observing the changes in UCL spectra caused by variations in pump power. The time-resolved fluorescence analysis examined the green and red light emission. The results indicated that the time decay increases as the dopant concentration increases, indicating the presence of non-radiative transitions at higher dopant concentrations. The FIR approach quantified the absolute sensitivity, yielding a result of 0.91% K^{-1} at a temperature of 573 K. These enhancements imply that the created material may be beneficial in developing sophisticated optoelectronic devices for versatile applications.

Acknowledgment

The authors thank the Functional Material Research Lab (FMRL) of the Delhi Technological University for all the essential research materials. Also, one of the authors thanks CSIR (08/0133(15560)) for their generous financial assistance.

Credit Authorship Contribution Statement

Ankita Banwal: Conceptualization, data curation, writing-original draft, made all the measurements. **Bhavya Kumar:** Writing, editing, and data curation. **Manoj Verma:** XRD analysis. **Ankur Shandilya:** Time decay measurements. **Bharti Singh:** Data curation. **Renuka Bokolia:** Supervision, reviewing, and editing of the original draft.

Data Availability

The above-mentioned authors have all the relevant data associated with this research work and will be dedicated to sharing them if they are asked to do so in the future.

References

[1] Z. Tianmin, Z. Yanqiu, W. Zhongli, and C. Baojiu, "Concentration effect and temperature quenching of upconversion luminescence in $BaGd_2ZnO_5:Er^{3+}/Yb^{3+}$ phosphor," *J. Rare Earth.*, vol. 33, no. 7, pp. 686, 2015, doi: 10.1016/S1002-0721(14)60471-3.

[2] M. K. Mahata, T. Koppe, T. Mondal, C. B. Sewitz, K. Kumar, V. K. Rai, H. Hofsass, and U. Vetter, "Incorporation of Zn^{2+} ions into $BaTiO_3:Er^{3+}/Yb^{3+}$ nanophosphor: an effective way to enhance upconversion, defect luminescence, and temperature sensing," *Phys. Chem. Chem. Phys.*, vol. 17, pp. 20741-20753, 2015, doi: 10.1039/c5cp01874a.

[3] Q. Cao, D. Peng, H. Zou, J. Li, X. Wang, and X. Yao, "Up-conversion luminescence of Er^{3+} and Yb^{3+} co-doped $CaBi_2Ta_2O_9$ multifunctional ferroelectrics," *J. Adv. Dielectr.*, vol. 4, no. 03, pp. 1450018, 2014, doi: 10.1142/s2010135x14500180.

[4] M. Varshney, S. Soni, A. Banwal, M. Narwan, M. Verma, and R. Bokolia, "Effect of Er^{3+} ion incorporation on the structural, photoluminescence, and ferroelectric properties of $K_{0.5}Na_{0.5}NbO_3$ ceramic for optoelectronic applications," *Appl. Phys. A*, vol. 130, pp. 267, 2024, doi: 10.1007/s00339-024-07447-1.

[5] A. Banwal and R. Bokolia, "Phase evolution and microstructure of $BaBi_2Nb_2O_9$ ferroelectric ceramics," *Mater. Today Proc.*, vol. 46, part 20, pp. 10121-10124, 2021, doi: 10.1016/j.matpr.2020.09.380.

Journal Pre-proof

[6] R. D. Shannon, "Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides," *J. Acta Cryst.*, A32, pp. 751–767, 2016, doi: 10.1107/S0567739476001551.

[7] A. Banwal and R. Bokolia, "Effect of Er^{3+} ion doping on structural, ferroelectric and up/down conversion luminescence in $BaBi_2Nb_2O_9$ ceramic," *Mater. Today Proc.*, vol. 47, part 14, pp. 4692-4695, 2021, doi: 10.1016/j.matpr.2021.05.545.

[8] M. X. Façanha, J. P. C. do Nascimento, M. A. S. Silva, M. C. C. Filho, A. N. L Marques, A. G. Pinheiro, and A. S. B. Sombra, "Up-conversion emission of Er^{3+}/Yb^{3+} co-doped $BaBi_2Nb_2O_9$ (BBN) phosphors," *J. Lumin.*, vol. 183, pp. 102-107, 2017, doi: 10.1016/j.jlumin.2016.08.011.

[9] Z. Chen, L. Sheng, X. Li, P. Zheng, W. Bai, L. Li, F. Wen, W. Wu, L. Zheng, and J. Cui, "Enhanced piezoelectric properties and electrical resistivity in W/Cr co-doped $CaBi_2Nb_2O_9$ high-temperature piezoelectric ceramics," *Ceram. Int.*, vol. 45, no. 5, pp. 6004–6011, 2019. doi: 10.1016/j.ceramint.2018.11.252.

[10] Z. Yao, R. Chu, Z. Xu, J. Hao, W. Li, and G. Li, "Processing and enhanced electrical properties of $Sr_{1-x}(K_{0.5}Bi_{0.5})_xBi_2Nb_2O_9$ lead-free piezoelectric ceramics," *Ceram. Int.*, vol. 42, no. 9, pp. 10619-10623, 2016, doi: 10.1016/j.ceramint.2016.03.156.

[11] X. Hui, D. Peng, H. Zou, J. Li, Q. Cao, Y. Li, X. Wang, and X. Yao, "A new multifunctional Aurivillius oxide $Na_{0.5}Er_{0.5}Bi_4Ti_4O_{15}$: Up-conversion luminescent, dielectric, and piezoelectric properties," *Ceram. Int.*, vol. 40, part 4A, pp. 12477-12483, 2014, doi: 10.1016/j.ceramint.2014.04.102.

[12] R. Ramaraghavulu and S. Buddhudu, "Structural and dielectric properties of $BaBi_2Nb_2O_9$ ferroelectric ceramic powders by a solid state reaction method," *Ferroelectrics*, vol. 460, no. 1, pp. 57–67, 2014. doi: 10.1080/00150193.2014.874924

[13] F. Luo, J. Xing, Y. Qin, Y. Zhong, F. Shang, and G. Chen, "Up-conversion luminescence, temperature sensitive and energy storage performance of lead-free transparent Yb^{3+}/Er^{3+} co-doped $Ba_2NaNb_5O_{15}$ glass-ceramics," *J. Alloy. Compd.*, vol. 910, pp. 164859, 2022, doi: 10.1016/j.jallcom.2022.164859.

[14] A. Pandey, S. Som, V. Kumar, V. Kumar, K. Kumar, V. K. Rai, and H.C. Swart, "Enhanced upconversion and temperature sensing study of Er^{3+}–Yb^{3+} co-doped tungsten–tellurite glass," *Sens. and Act. B: Chem.*, vol. 202, pp. 1305-1312, 2014, doi: 10.1016/j.snb.2014.06.074.

[15] R. Swami, A. Shandilya, A.K. Shukla, P. S. Kumar, and K. Sreenivas, "Luminescence intensity ratio based wide range contactless temperature sensing properties of Er^{3+}/Yb^{3+} ions co-doped $La_2Ti_2O_7$ ceramics," *J. Alloy. Compd.*, vol. 948, pp. 169610, 2023, doi: 10.1016/j.jallcom.2023.169610.

[16] P. Wang, X. Wang, G. Li, Y. Li, X. Yao, and Z. Pan, "Energy density capability and upconversion luminescence in Er^{3+}/Yb^{3+}-codoping BNT-based ferroelectric thin films," *Ceram. Int.*, vol. 48(19), pp. 28606-28613, 2022, doi: 10.1016/j.ceramint.2022.06.174.

[17] Z. Zhou, X. Dong, and H. Chen, "Structural and electrical properties of W^{6+} doped Bi_3TiNbO_9 high-temperature piezoceramics," *J. Am. Ceram. Soc.*, vol. 89 (5), pp. 1756-1760, 2006, doi: 10.1111/j.1551-2916.2006.00957.x.

[18] Q. Hou, B. Yang, C. Ma, Z. Zhou, R. Liang, H. Li, and X. Dong, "Tailoring structure and piezoelectric properties of $CaBi_2Nb_2O_9$ ceramics by W^{6+} doping, " *Ceram. Int.*, vol. 48, pp. 16677-16684, 2022, doi: 1 10.1016/j.ceramint.2022.02.216.

[19] A. Pandey and V. K. Rai, "Improved luminescence and temperature sensing performance of Ho3+–Yb3+–Zn2+:Y2O3 phosphor," *Dalton Trans.*, vol. 42, 11005, 2013, doi: 10.1039/c3dt50592h.

[20] R. Bokolia, M. Mondal, V. K. Rai, and K. Sreenivas, "Enhanced infrared-to-visible up-conversion emission and temperature sensitivity in (Er^{3+}, Yb^{3+}, and W^{6+}) tri-doped $Bi_4Ti_3O_{12}$ ferroelectric oxide," *J. Appl. Phys.*, vol. 121, no. 8, 2017, doi: 10.1063/1.4977006.

[21] A. Banwal and R. Bokolia, "Enhanced upconversion luminescence and optical temperature sensing performance in Er^{3+} doped $BaBi_2Nb_2O_9$ ferroelectric ceramic," *Ceram. Int.*, vol. 48, no. 2, pp. 2230-2240, 2022, doi: 10.1016/j.ceramint.2021.09.314.

[22] A. Banwal and R. Bokolia, "Efficient tunable temperature sensitivity in thermally coupled levels of Er^{3+}/Yb^{3+} co-doped BaBi2Nb2O9 ferroelectric ceramic," *J. Lumin.*, vol. 263, pp. 120071, 2023, doi: 10.1016/j.jlumin.2023.120071.

[23] A. Banwal, M. Verma, B. Singh, and R. Bokolia, "Temperature stability and improved energy storage efficiency of $BaBi_2Nb_2O_9$: Er^{3+}/Yb^{3+} relaxor ferroelectric ceramic under moderate electric fields" *Appl. Phys. A*, vol. 130, pp. 334, 2024, doi: 10.1007/s00339-024-07475-x

[24] P. Du and J. S. Yun, "Effect of molybdenum on upconversion emission and temperature sensing properties in $Na_{0.5}Bi_{0.5}TiO_3$: Er/Yb ceramics," *Ceram. Int.*, vol. 41, pp. 6710–6714, 2015, doi: 10.1016/j.ceramint.2015.01.113.

[25] A. Banwal and R. Bokolia, "Thermometric sensing performance in Erbium modified $SrBi_{2-x}Nb_2Er_xO_9$ ferroelectric ceramic for optoelectronic devices," *Ceram. Int.*, vol. 48 part 23A, pp. 34405-34414, 2022, doi: 10.1016/j.ceramint.2022.08.019.

[26] Z. Peng, L. Chen, Y. Xiang, and F. Cao, "Microstructure and Electrical Properties of Lanthanides-doped $CaBi_2Nb_2O_9$ Ceramics," *Mater. Res. Bull.*, vol. 148, 111670, 2022. 10.1016/j.materresbull.2021.111670.

[27] A. Shandilya, R. S. Yadav, A. K. Gupta, and K. Sreenivas, "Effects of Yb^{3+} ion doping on lattice distortion, optical absorption and light upconversion in Er^{3+}/Yb^{3+} co-doped $SrMoO_4$ ceramics," *Mater. Chem. Phys.*, vol. 264, 124441, 2021. doi: 10.1016/j.matchemphys.2021.124441.

[28] P. Li, Z. Zhang, Z. Xu, H. Sun, Q. Zhang, and X. Hao, "Photomodulated cryogenic temperature sensing through a photochromic reaction in $Na_{0.5}Bi_{2.5}Ta_2O_9$: Er/Yb multicolour upconversion," *Opt Express.*, vol. 31(5), pp. 7047–7059, 2023, doi: 10.1364/OE.469538.

[29] R. Bokolia, O. P. Thakur, V. K. Rai, S. K. Sharma, and K. Sreenivas, "Electrical properties and light up conversion effects in $Bi_{3.79}Er_{0.03}Yb_{0.18}Ti_{3-x}W_xO_{12}$ ferroelectric ceramics," *Ceram. Int.*, vol. 42, no. 5, pp. 5718–5730, 2016. doi: 10.1016/j.ceramint.2015.12.103.

[30] W. Ma, Y. Zhu, M. A. Marwat, P. Fan, B. Xie, D. Salamon, Z. G. Ye, and H. Zhang, "Enhanced energy-storage performance with excellent stability under low electric fields in BNT-ST relaxor ferroelectric ceramics," *J. Mater. Chem.* C, vol. 7, no. 2, pp. 281–288, 2019. doi: 10.1039/c8tc04447c.

[31] M. Narwan, A. Banwal, R. Sharma, and R. Bokolia, "Non-invasive thermal sensing and improved recoverable energy storage density of $Bi_{0.5}Na_{0.5}TiO_3$: Er^{3+} doped multifunctional ferroelectric ceramic," *J. Lumin.*, vol. 265, pp. 120236, 2024. doi: 10.1016/j.jlumin.2023.120236.

Journal Pre-proof

Highlights:

1. A detailed study of Er/Yb/W tri-doped $BaBi_{2-0.04-0.10}Nb_{2-z}Er_{0.04}Yb_{0.10}W_zO_9$ (BBN) system, and a comparison of undoped, Er-doped, Er/Yb co-doped and Er/Yb/W tri-doped BBN systems are presented.
2. The UCL spectra are traced at 980 nm and exhibit two distinctive green bands and one red band at 535 nm, 556 nm, and 673 nm, respectively.
3. The time-resolved fluorescence analysis was done for green and red emissions, and it was observed that the time decay increased for higher dopant concentrations.
4. The FIR technique measured the absolute sensitivity and had a value of 0.91% K^{-1} at 573 K.

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575090

Improving Intrusion Detection in a Software Defined Network using Hybrid CNN and Bi-LSTM

Ashutosh Bhattarai
Department of Information Technology
Delhi Technological University
New Delhi, India
ashutoshbhattarai2058@gmail.com

Unmik Gyawali
Department of Information Technology
Delhi Technological University
New Delhi, India
unmikgyawali@gmail.com

Abhishek Verma
Department of Information Technology
Delhi Technological University
New Delhi, India
abhishekcms08@gmail.com

Virender Ranga
Department of Information Technology
Delhi Technological University
New Delhi, India
virenderranga@dtu.ac.in

***Abstract*—A Software Defined Networking is a modern offering against traditional network systems to make network control more efficient by reducing the dependency on the hardware side. With a logical control plane, it is essential to distinguish between benign and malicious data traffic. We explore the possibility of a hybrid neural network architecture for intrusion detection and traffic monitoring in an SDN environment. The suggested architecture stacks a Convolutional Neural Network (CNN) with Bidirectional Long Short Term Recurrent Neural Networks (BiLSTM) to provide a specialized architecture for distinguishing the incoming network traffic patterns in SDN. The results emphasize the hybrid architecture's effectiveness in detecting spatial and temporal features, contributing to improved network security in SDN environments.**

***Keywords*—Benign, Malicious, SDN, CNN, BiLSTM**

I. Introduction

Software Defined Networking is a modern approach that promises to change the network performance by splitting the network control from hardware devices like routers, switches, etc [1]. Unlike traditional approaches, SDN completely isolates the application tier from the control tier. Software defined networking reduces dependency on the hardware side by the logical centralisation of the control plane. However, it exposes the network to cyber threats. Despite available tools and techniques for enhancing security and detecting abnormal online behaviour, attackers persist in breaching confidentiality and compromising data. With more dependency on the software side, any intrusion into the control tier is a danger. So, there is a need for creating a robust solution for network traffic classification to boost accuracy and adaptability in network security. Machine learning based intrusion detection systems have gained importance for adapting to this evolving SDN network. Machine learning models automate real time threat detection by analysing extensive network data and traffic patterns.

Intrusion attacks involve penetrating the control tier that controls the entire network architecture. The main attacks might be ARP Spoofing, ARP Flooding Attacks, and Traffic Sniffing [2]. Attacks like SYN Flooding are also responsible for sending continuous requests to the network, causing it to flood with requests, ultimately disrupting the network functionality. In a malicious ARP request, the header of the data is tampered with in such as way that the data is received by the attacker instead of the genuine user for whom the data was originally intended. Once exploiters gain access to the control tier, they can easily manipulate the entire network, add their devices and steal valuable information from the user.

The primary motivation for this work is the understanding that traditional network security mechanisms often fall short in detecting network intrusion. The proposed approach uses the combination of CNN and BiLSTM to automatically extract and evaluate intricate elements in data packets, helping to achieve distinction of network traffic.

II. Background

A. Convolutinal Neural Networks

Operating under supervised learning, Convolutional Neural Network (CNN) is an essential model in computer vision. It makes good use of the weight sharing idea to solve the parameter explosion issue. A CNN architecture comprises of one or more convolutional layers, pooling layers, and fully connected layers. CNNs are widely employed in facial recognition and image processing applications. In the convolutional layer, linear operations are carried out by a filter that travels over the output of the previous layer [3]. The RELU activation function returns null for negative values in the feature map but returns the value for positive values. Multiple convolutional layers are used in CNNs, the first few layers extract simple characteristics like corners and edges, whereas the latter layers capture more intricate information. By shrinking the window size, the pooling layer lowers computation expenses. Given the wide deployment of CNN in various sub domains of deep

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:33:38 UTC from IEEE Xplore. Restrictions apply.

learning, it is worth analyzing if the inclusion of a CNN component can help enhance network intrusion detection [4].

B. Bidirectional Long Short Term Memory

While processing sequential data step by step, Recurrent Neural Networks (RNN) encounter obstacles such as exploding gradients. Problems in RNNs are addressed via memory units and gate mechanisms in LSTMs. Unlike LSTM, which processes input sequences in a unidirectional manner, Bidirectional Long Short-Term Memory (BiLSTM) consists of gates to allow information to flow in both directions, making it possible to efficiently capture contextual information. It uses two hidden states to store information about both the present and the future context.

In a BiLSTM architecture, the output of the first layer is the input to the second layer. The output of the second layer is the concatenation of the final unit outputs from both the forward and backward directions of the first layer [5]. Because BiLSTM can describe bidirectional connections in complicated sequential data, it is useful for tasks like natural language processing, gesture recognition, and time-series prediction [6]. So, they may prove useful in dealing with sequential network traffic data.

C. The proposed CNN-BiLSTM Hybrid Architecture

With the integration of both BiLSTM and CNN in a hybrid architecture, the suggested model improves traffic data feature extraction and pattern recognition in software-defined networking. The architecture starts with a CNN component. The convolutional layer applies a set of filters to the incoming network traffic to extract spatial data and look for local patterns within the traffic. CNNs excel when it comes to pattern recognition, so they can prove efficient in extracting malicious features from the network data packets. Because they are capable of detecting a pattern regardless of its position in the input sequence, CNNs can be the right choice for dealing with SDN network traffic data. Following the convolutional operation, the ReLU activation function is applied to introduce non-linearity into the network [7].

The dual bidirectional LSTM layers then capture both the forward and the backward temporal features. The first layer captures the immediate dependencies of the given source while the second layer looks for long-term dependencies within the source. Because BiLSTMs can handle sequences of variable length without the need for padding, they can be effective in dealing with network traffic data, which mostly come in variable length sequences. Furthermore, both LSTMs and BiLSTMs are known to be capable in learning from corrupted information, so they may be efficient in extracting meaningful representations from noisy input network traffic. This kind of special architecture may allow the model to effectively infer the sequential nature of the SDN traffic.

After the BiLSTM layers, a dense layer with 3 units along with a softmax activation function is introduced and finally, probabilities for each intrusion type is computed.

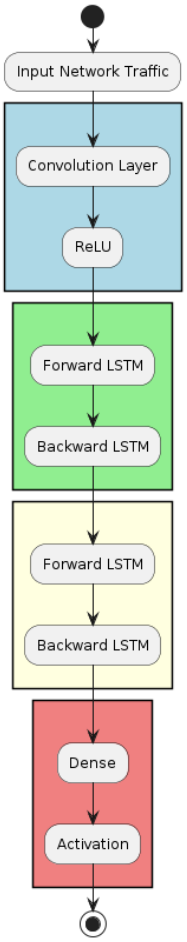


Fig. 1. Framework for the proposed architecture

The rest of the article proceeds as the following: Section III provides a brief summary of previous related works, Section IV lists out the methodology and the model workflow, Section V presents the experimental analysis and Section VI lays out the conclusions.

III. Literature Review

For the last few years, several studies have been conducted to improve intrusion detection in an SDN environment [8]. Given the growing need for SDN, it is essential to come up with ways to protect the network. Alzahrani and Alenazi first explored the potential of machine learning algorithms, such as Decision Tree, Random Forest, and XGBoost, for intrusion detection, avoidance, and traffic monitoring in Software Defined Networking [9]. The study primarily focused on detecting and classifying various types of common network attacks, like the Distributed Denial of Service (DDoS). The authors used the NSL-KDD dataset [10] for model training and achieved an impressive accuracy of approximately 95.95%.

In recent research conducted on intrusion detection in an SDN environment, deep learning techniques have been the most preferred choice, particularly CNN and LSTM models

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:33:38 UTC from IEEE Xplore. Restrictions apply.

[11]. These Deep learning models have been explored more as they have proven efficient in capturing temporal patterns as well as contextual information based on any sequence within the training dataset.

Tang, T. A., et al. implemented a model focused on detecting and identifying flow-based anomalies in an SDN environment [12]. Only six characteristics were focused out of all from the NSL-KDD dataset to reduce the overall computational cost. Comprising 12 neurons in the first layer, 6 neurons in the second layer, and 3 neurons in the final layer [13], the proposed three hidden layered Deep Neural Network model achieved an accuracy of about 75%. The score was considered too poor to be adopted for real-world scenarios. Thus for enhancement, the authors proposed a Gated Recurrent Unit (GRU), which increased the accuracy to about 89%.

Elsayed et al. proposed that having a memory understanding of the attack pattern could enhance intrusion detection. This could be done by utilizing the capability of long short-term memory with the understanding of a convolutional neural network [14]. This aggregation enabled the model to capture both temporal as well as spatial features effectively. To address overfitting, the authors employed the following regularization techniques: The L2 norm and Dropout. The authors tested the model's capability against the publicly available InSDN dataset [15]. The model gained an impressive accuracy of around 96%, which is greater than the accuracy of CNN and LSTM implemented individually.

Jiang et al. suggested a Deep Hybrid Intrusion Detection Technique, which used the combined capabilities of both convolutional neural networks and bidirectional long short-term memory networks [16]. To mitigate noise and establish a balanced dataset, the authors introduced One-Sided Selection (OSS). The model was tested against the following publicly available datasets: NSL-KDD, and UNSW-NB15. The results demonstrated promising performance, achieving an accuracy of around 83%.

Khan et al. introduced a Two-Stage Deep Learning Architecture to address the demand for efficient network intrusion detection [17]. The model comprises two decisional stages, the first differentiates the incoming network activity into normal and abnormal classes. For each instance, a score is assigned. This assigned score is then used in the latter decision stage, which distinguishes between normal network state and various attack classes. The proposed model was validated by conducting experiments with the KDD99 and the UNSW-NB15 datasets. While the model did achieve an impressive accuracy rate of around 99% for the KDD99 dataset and about 89% for the NSW-NB15 dataset, it is essential to highlight that networking-specific intrusion detection was not the focus point of the study.

ARP spoofing and flooding present significant security risks in network environments. ARP attacks, which include common attacks like DoS (Denial of Service), MITM (Man in the Middle), and host impersonation pose severe security risks to any network. Throughout the years, various diverse solutions have been proposed to detect and mitigate ARP attacks. Meghana, J. S. et al. highlighted the effectiveness of SDNs in reducing ARP attacks [18]. The authors proposed SDN-based techniques, employing buffer management, Bayes-based algorithms, and Network Flow Guard modules to enhance ARP attack detection. The study highlights the benefits of SDN features, such as centralized control and dynamic policy enforcement.

All these studies collectively underscore the efficiency and potential of deep neural networks in SDN security. So, it is worth studying if the proposed neural network architecture can give good results in SDN intrusion detection.

IV. Methodology

The workflow is as follows : Network traffic data simulated for the SDN environment is collected from Mendeley data [19], [20]. Preprocessing activities are carried out and the data is subjected to model training. Post training, performance metrics are calculated and results are compared with other neural network models.

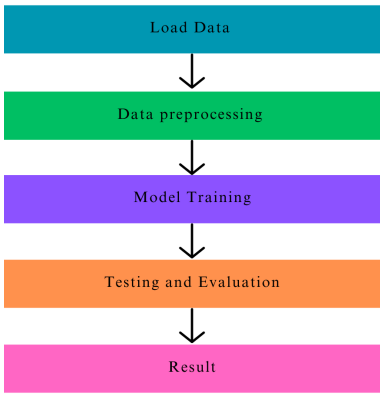


Fig. 2. Proposed workflow

A. Dataset

Many existing SDN datasets fail to provide diversity and relevance to current network conditions. The present research is limited to very few attacks and fails to cover ARP attacks. The datasets also lack genuine network traffic and consist of redundant and null values. To address this problem, the dataset chosen is generated using the mininet emulator, offering a realistic environment for experimental purposes. The simulated network consists of a Ryu controller, interconnected with 27 hosts and 12 switches, and uses a tree architecture with certain fan-out and depth parameters. The simulation examines network traffic that is both benign and malicious, focusing on ARP poison and ARP flood attacks.

The last column in the dataset acts as a class label, designating malicious (labeled 1 or 2) and legitimate (labeled 0) traffic. A total of 104,345 rows of data are gathered throughout a 250-minute simulation, and another 13,000 rows are obtained over a three-hour simulation. For model assessment, the used

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:33:38 UTC from IEEE Xplore. Restrictions apply.

dataset is subsequently divided into training and validation, where the former is used to train the model whereas the latter stage is essential for determining the model's ability to handle unknown input outside of its training set.

B. Data Preprocessing

Several data preprocessing techniques have been applied to ensure that the dataset chosen is clean and suitable for classification. Missing values are addressed by removing all the rows containing any missing values. Duplicate rows are eliminated to reduce redundancy and feature selection is performed by dropping columns that are irrelevant for model training. All these preprocessing activities enhance the quality of the chosen dataset for the subsequent model building.

Feature scaling is used to standardize the feature variables. Feature scaling provides a guarantee that all of the features are on a comparable scale, preventing some features from outperforming others throughout the training step. Because the network traffic is generated across an extended time period, there is a large scale in between certain features, such as the flow of packets. Thus, these features are normalised to have a range in between 0 to 1. The mathematical equation for the normalisation of the feature values is as follows :

$$f_{\text{normalised}} = \frac{f - \text{minimum}(f)}{\text{maximum}(f) - \text{minimum}(f)} \tag{1}$$

where $f_{\text{normalised}}$ is the normalised feature value, f is the original feature value, $\text{minimum}(f)$ is the minimum feature value in the dataset, and $\text{maximum}(f)$ is the maximum feature value in the dataset. Additionally, the categorical labels are transformed into binary vectors using one-hot encoding, thus enabling the model to learn from the categorical labels appropriately.

C. Model Training

To address issues like overfitting or underfitting and enhance overall model performance, it is essential to conduct hyperparameter adjustments. In this regard, a randomized search approach is employed to systematically define hyperparameters, aiming to optimize the model's accuracy and mitigate issues related to overfitting or underfitting. The hyperparameter configuration for the proposed model is as follows:

TABLE I
HYPERPARAMETER CONFIGURATION

Attributes	Values
Kernel Size	5
Filters	64
BiLSTM Layer	64, 128
Activation Function	ReLU
Kernel Regularization	L2 (0.01)
Loss Function	Categorical Cross Entropy
Optimizer	Adam
Dropout Rate	0.06

To evaluate the outcomes of the proposed model, the following neural network models are also trained on the same dataset: Vanilla RNN, BiLSTM, CNN, LSTM [21].

V. EXPERIMENTAL ANALYSIS

A. Evaluation Metrics

After training, to assess the performance of the model, the following evaluation metrics are computed:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

The performance metrics shown in Equations (1), (2), and (3) above are derived from the four quadrants of the confusion matrix. The four quadrants are True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). A model correctly predicting a positive instance is known as a true positive. Similarly, when the model correctly predicts a negative instance, it is known as a true negative. When the model incorrectly predicts a positive outcome for a negative instance then it is known as false positive. Similarly, when the model incorrectly predicts a negative outcome for a positive instance it is known as false negative.

The Area Under the Receiver Operator Characteristics Curve (AUC-ROC) indicator measures the model's capability to differ across various categories. AUC, which is computed through the ROC curve, shows the sensitivity against the false positive rate. The classifier's effectiveness in distinguishing between network traffic being either benign or malicious can be enhanced with a higher AUC-ROC value, which should be positive and less than one.

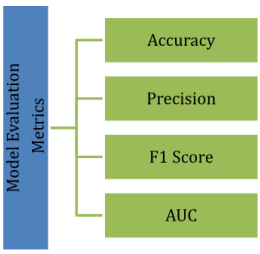


Fig. 3. Model evaluation metrics

Intrusion Detection Rate calculates the ratio of actual intrusions that are correctly identified by the model. For a particular class i, the intrusion detection rate can be calculated as the following:

$$IDR_i = \frac{TP_i}{\text{Z}i} \tag{5}$$

where IDR_i is the Intrusion Detection Rate for class i, TP_i is the total number of true positives for class i and Z is the total instances for the class i.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:33:38 UTC from IEEE Xplore. Restrictions apply.

B. Results and Discussion

The calculated evaluation metrics for the CNN-BiLSTM model, along with other neural network models are as follows:

TABLE II
EVALUATION SCORES

Model	Acc. (%)	Prec. (%)	F1-Score	AUC ROC
CNN	84.48	95.6	0.90	0.98
RNN	86.99	95.7	0.89	0.98
LSTM	87.57	95.8	0.90	0.98
BiLSTM	88.67	95.6	0.90	0.98
BiLSTM-CNN	**96.01**	**96.0**	**0.96**	**0.99**

With a high accuracy result, the model shows great promise in accurately classifying network traffic and distinguishing between normal and malicious data. The high precision score shows the model's effectiveness in minimizing false positives, which is crucial for mitigating operational disruptions and false alarms in SDN deployment. With a similar recall score, the model is able to identify a substantial number of true positive cases. Additionally, the high F1 score indicates that there is a perfect balance between precision and recall.

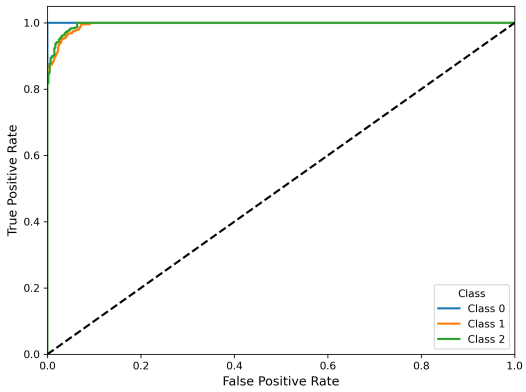


Fig. 4. Receiver Operating Characteristics (ROC) curve

Talking about other models, BiLSTM achieves a respectable accuracy of 88.67. Other traditional models also demonstrate competitive accuracies but lack the comprehensive ability of the hybrid model to encompass both local and global dependencies simultaneously. The ability of the combined architecture to extract both sequential and spatial feature extraction makes it a more suitable choice than the other baseline models.

TABLE III
INTRUSION DETECTION RATES

Class	Intrusion Detection Rate (%)
ARP Poison Traffic (Class 1)	92.44
ARP Flood Traffic (Class 2)	95.59

The Intrusion Detection Rate (IDR) measures the model's ability to correctly detect the instances of intrusion for each class. About 93 % of all the instances classified as ARP Poison Traffic were correctly identified. Similarly, around 96 % of all the instances classified as ARP Flood Traffic were correctly identified. These high detection rates further ascertain the architecture's effectiveness in identifying the instances of the corresponding intrusion classes.

While the model shows strong performance across the chosen evaluation metrics, there is still opportunity for further optimisation of the model. In the future, several strategies like fine tuning the hyper parameters, optimizing preprocessing activities and integrating ensemble learning techniques can be explored to enhance the precision. The proposed model's deviation from traditional architectures may be influenced by unique aspects of SDN traffic patterns, warranting further investigation into the dataset's intricacies.

VI. CONCLUSION

In this study, we proposed and evaluated a hybrid neural network architecture to classify ARP attack data in a software defined networking environment. To compare how well the model performs, other traditional neural network architectures were also trained on the same dataset. Each architecture's accuracy, precision, F1 score and AUC-ROC scores have been calculated and analyzed. The experimental analysis shows promising results for all the neural network architectures. But, on average of all performance metrics, the proposed BiLSTM-CNN hybrid model performed admirably, demonstrating significant promise and effectiveness. The model performed well in identifying occurrences, thanks to its comprehensive methodology, as demonstrated by the encouraging findings.

Identifying harmful incoming network traffic is crucial for delivering timely countermeasures against cyber threats. Future work will prioritize the development and implementation of a custom SDN infrastructure, enabling real-time data collection to assess and refine the model's functionality in practical scenarios. Comprehensive evaluations of the model's latency and throughput on live SDN networks will be conducted to gauge its real-world performance accurately. The performance of the model can also be assessed in other wider domains in the future [22]–[27].

REFERENCES

[1] A. Ominike, E. Seun, A. O., and F. Osisanwo, "Introduction to Software Defined Networks (SDN)," *International Journal of Applied Information Systems*, vol. 11, pp. 10–14, Dec. 2016.
[2] A. Alhaj and N. Dutta, "Analysis of Security Attacks in SDN Network: A Comprehensive Survey," pp. 27–37, Jan. 2022.
[3] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," *ArXiv e-prints*, Nov. 2015.
[4] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects.," *IEEE transactions on neural networks and learning systems*, vol. 33, pp. 6999–7019, Dec. 2022. Place: United States.
[5] S. Sinha and D. V. Le, "Completely automated cnn architecture design based on vgg blocks for fingerprinting localisation," in *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, IEEE, 2021.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:33:38 UTC from IEEE Xplore. Restrictions apply.

[6] M. Rao, Q. Li, D. Wei, and M. J. Zuo, "A deep bi-directional long short-term memory model for automatic rotating speed extraction from raw vibration signals," *Measurement*, vol. 158, p. 107719, 2020.
[7] R. Raut and D. Goel, "Mood-based emotional analysis for music recommendation." EasyChair Preprint no. 10377, EasyChair, 2023.
[8] Z. Mustafa, R. Amin, H. Aldabbas, and N. Ahmed, "Intrusion detection systems for software-defined networks: a comprehensive study on machine learning-based techniques," *Cluster Computing*, pp. 1–27, 2024.
[9] A. O. Alzahrani and M. J. F. Alenazi, "Designing a Network Intrusion Detection System Based on Machine Learning for Software Defined Networks," *Future Internet*, vol. 13, no. 5, 2021.
[10] L. Dhanabal and S. Shantharajah, "A study on nsl-kdd dataset for intrusion detection system based on classification algorithms," *International journal of advanced research in computer and communication engineering*, vol. 4, no. 6, pp. 446–452, 2015.
[11] R. Chaganti, W. Suliman, V. Ravi, and A. Dua, "Deep Learning Approach for SDN-Enabled Intrusion Detection System in IoT Networks," *Information*, vol. 14, no. 1, 2023.
[12] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for Network Intrusion Detection in Software Defined Networking," in *2016 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, pp. 258–263, 2016.
[13] T. Pattarabanjird, C. Cress, A. Nguyen, A. Taylor, S. Bekiranov, and C. McNamara, "A machine learning model utilizing a novel snp shows enhanced prediction of coronary artery disease severity," *Genes*, vol. 11, no. 12, 2020.
[14] M. S. ElSayed, N.-A. Le-Khac, M. A. Albahar, and A. Jurcut, "A novel hybrid model for intrusion detection systems in SDNs based on CNN and a new regularization technique," *Journal of Network and Computer Applications*, vol. 191, p. 103160, 2021.
[15] M. Elsayed, N.-A. Le-Khac, and A. Jurcut, "Insdn: A novel sdn intrusion dataset," *IEEE Access*, 09 2020.
[16] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network Intrusion Detection Combined Hybrid Sampling With Deep Hierarchical Network," *IEEE Access*, vol. 8, pp. 32464–32476, 2020.
[17] Z. Ahmad, A. Shahid Khan, C. Shiang, and F. Ahmad, "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Transactions on Emerging Telecommunications Technologies*, vol. 32, Jan. 2021.
[18] J. S. Meghana, T. Subashri, and K. Vimal, "A survey on ARP cache poisoning and techniques for detection and mitigation," in *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, pp. 1–6, 2017.
[19] Nisha Ahuja, "ARP Poisoning and Flood attack in SDN," 2022.
[20] N. Ahuja, G. Singal, D. Mukhopadhyay, and A. Nehra, "Ascertain the efficient machine learning approach to detect different ARP attacks," *Computers and Electrical Engineering*, vol. 99, p. 107757, 2022.
[21] M. S. E. Sayed, N.-A. Le-Khac, M. A. Azer, and A. D. Jurcut, "A flow-based anomaly detection approach with feature selection method against ddos attacks in sdns," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 4, pp. 1862–1880, 2022.
[22] A. Verma, V. Ranga, and D. K. Vishwakarma, "Forecasting of satellite based carbon-monoxide time-series data using a deep learning approach," in *2023 4th International Conference on Innovative Trends in Information Technology (ICITIIT)*, pp. 1–4, IEEE, 2023.
[23] A. Verma, V. Ranga, and D. K. Vishwakarma, "A novel approach for forecasting pm2. 5 pollution in delhi using catalyst," *Environmental Monitoring and Assessment*, vol. 195, no. 12, p. 1457, 2023.
[24] A. Verma, V. Ranga, and D. K. Vishwakarma, "Combating respiratory health issues with intelligent no2 level prediction from sentinel 5p satellite," in *2023 IEEE 20th India Council International Conference (INDICON)*, pp. 882–886, IEEE, 2023.
[25] A. Verma, V. Ranga, and D. K. Vishwakarma, "Breath-net: a novel deep learning framework for no2 prediction using bi-directional encoder with transformer," *Environmental Monitoring and Assessment*, vol. 196, no. 4, p. 340, 2024.
[26] A. Yadav and D. K. Vishwakarma, "Mrt-net: Auto-adaptive weighting of manipulation residuals and texture clues for face manipulation detection," *Expert Systems with Applications*, vol. 232, p. 120898, 2023.
[27] K. Tripathi and A. Verma, "Remote sensing and machine learning fusion: A robust framework for land use and land cover change detection," in *2024 IEEE International Conference for Women in Innovation, Technology Entrepreneurship (ICWITE)*, pp. 434–439, 2024.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:33:38 UTC from IEEE Xplore. Restrictions apply.

NATURALISTA CAMPANO
ISSN: 1827-7160
Volume 28 Issue 1, 2024

Innovative Channel Estimation Method for Rayleigh, Rician, and Nakagami Channels: The Modified M-Estimator-I

Prakhar Hari[1], Sanjay Mathur[2], Prachi Mishra[3], Pooja Singh[4], Hari Kumar Singh[5*]

[1]Department of Electronics and Communication Engineering, G B P U A T Pantnagar
[2]Department of Electronics and Communication Engineering, G B P U A T Pantnagar
[3]Department of Electronics and Communication Engineering, G B P U A T Pantnagar
[4]Department of Electrical Engineering, Delhi Technological University, Delhi
[5*]Department of Electronics and Communication Engineering, M J P Rohilkhand University Bareilly

Email: [1]prakharhari6766@gmail.com, [2]mathur_sanjay@rediffmail.com, [3]mishraprachiece@gmail.com, [4]er.poojadc11@gmail.com

Corresponding Author: [5*]hari.singh@mjpru.ac.in

Abstract: In modern wireless communication, achieving higher data rates, reliability, and spectral efficiency necessitates advanced transmission technologies. These technologies face challenges from unpredictable wireless propagation environments, particularly channel fading, which causes signal strength variations. Accurate channel estimation techniques are crucial for adapting transmission parameters to current channel conditions and optimizing performance. Orthogonal Frequency Division Multiplexing (OFDM) mitigates multipath propagation effects, and Multiple-Input Multiple-Output (MIMO) technology enhances capacity and reliability. The combined MIMO-OFDM framework relies on precise channel estimation, but traditional methods often fail in dynamic, complex channels. This research introduces a novel channel estimation technique, termed Modified M-estimator-I, which provides enhanced accuracy and resilience in estimating channel state information. Building on robust statistical estimation principles, the Modified M-estimator-I adapts to channel variations, offering more reliable performance across Rayleigh, Rician, and Nakagami fading environments. This innovative approach addresses the limitations of conventional estimators by incorporating advanced statistical methodologies to mitigate outliers and noise, thus optimizing the estimation process in diverse fading conditions.

Index: - Channel Estimation, M-estimator-I, Fading Channel, Pilot-symbol based Channel estimation

1. Introduction

In the realm of modern wireless communication, the quest for higher data rates, increased reliability, and enhanced spectral efficiency has driven significant advancements in transmission technologies. However, these advancements are continually challenged by the unpredictable and often adverse nature of wireless propagation environments. Channel fading, a phenomenon characterized by variations in signal strength over time and space, poses a formidable obstacle to the reliable transmission of data. To counteract these effects, accurate channel estimation techniques are indispensable, as they enable the adaptation of transmission parameters to the prevailing channel conditions, thereby optimizing performance.

Orthogonal Frequency Division Multiplexing (OFDM) has emerged as a cornerstone technology in contemporary wireless communication systems, renowned for its robustness against inter-symbol interference (ISI) and inter-carrier interference (ICI). By transmitting data across multiple orthogonal subcarriers, OFDM effectively mitigates the impact of multipath propagation. Transmitted symbols can be accurately recovered through effective channel estimation, which corrects the channel impulse response (CIR) and determines the channel's impact on the transmitted symbols. [1] MLE and MMSEE techniques for CIR estimation in OFDM systems was discussed. [2] The challenge of large CSI feedback in massive MIMO systems by proposing an efficient scheme using spatial correlation and compressed sensing (CS) techniques was discussed in [3]. Hence better reconstruction of transmitted symbols is required.

Complementing OFDM, Multiple-Input Multiple-Output (MIMO) technology leverages multiple transmitting and receiving antennas to enhance system capacity and reliability through spatial diversity and spatial multiplexing. The synergy between OFDM and MIMO, known as MIMO-OFDM, represents a powerful framework for high-performance wireless communication. Despite these technological advancements, the effectiveness of MIMO-OFDM systems is fundamentally dependent on precise channel estimation. Traditional channel estimation techniques, including pilot-based and blind estimation methods, have been widely studied and implemented. Various channel estimation techniques were studied for wireless communication system [4]. Least square channel estimation in MIMO-OFDM systems using block-type and comb-type pilot structures investigated in [5]. The estimation of channel parameters in OFDM and generalized frequency division multiplexing has been investigated in [6] and [7]. A combined approach of channel estimation and multi-user detection (MUD) in MIMO-OFDM systems to mitigate multi-access interference and reduce BER is proposed in [8]. LS and LMMSE-based channel techniques for channel parameters were compared in [9]. These conventional methods often fall short in addressing the dynamic and complex nature of wireless channels, particularly in the presence of severe fading. This necessitates the development of more sophisticated and adaptive estimation techniques. Robust multiuser detection using M-estimation in wireless communication systems was analysed in [10]. M-estimator-based approach compared with LS and MMSE approach in Rayleigh and Rician fading channel is done in [11]. Different channel estimation techniques are generally used for massive MIMO system, which are studied in [12].

In [13], a channel estimation technique for OFDM systems with fading channels is proposed. Rayleigh fading is commonly observed in environments with a high density of scatters, such as urban areas, where a direct line of sight between the transmitter and the receiver is absent. This results in a signal that is composed of numerous reflected and scattered paths, each with a different phase and amplitude. [14-15] proposes a method for analysing the performance of OFDM systems over Rayleigh fading channels, considering the effects of phase noise and carrier frequency offset. Rician fading, on the other hand, occurs in scenarios where there is a dominant line of sight path in addition to the scattered paths, leading to a more stable signal compared to Rayleigh fading. Nakagami fading provides a versatile model that can describe a wide range of fading conditions, from severe fading like Rayleigh to mild fading closer to Rician.

This research introduces a novel channel estimation technique, termed Modified M-estimator-I. The Modified M-estimator-I builds upon the principles of robust statistical estimation to provide enhanced accuracy and resilience in estimating channel state information. Unlike conventional estimators, which may be sensitive to outliers and noise, the Modified M-estimator-I incorporates advanced statistical methodologies to adapt to the variations in the channel conditions, offering a more reliable and robust performance across diverse fading environments.
In this paper, Section II outlines the system model. Section III explains the various channel estimation techniques and introduces the proposed Modified M-estimator-I. Section IV presents the simulation results, and Section V offers concluding remarks. The paper concludes with a list of references.

2. System MODEL

(A) OFDM and MIMO-OFDM System

Orthogonal Frequency Division Multiplexing (OFDM) is a wireless technology that transmits data on orthogonal subcarriers, allowing symbols to be closely packed without interference, thereby saving bandwidth. It uses Multi-Carrier Modulation to partition high-rate data into parallel bit streams, modulated by orthogonal subcarriers with IFFT and FFT for modulation and demodulation. Guard bands, cyclic prefixes, and pilot symbols are used to eliminate ISI, ICI, and assist in channel estimation. OFDM is robust against ISI, ICI, and fading, with enhanced performance when coupled with Multiple-Input Multiple-Output (MIMO) systems. MIMO employs multiple transmitting and receiving antennas to enhance system capacity and coverage through spatial multiplexing and diversity techniques. MIMO-OFDM improves link reliability and spectrum efficiency by increasing signal gain, reducing multi-path fading, and boosting channel capacity and SNR through spatial multiplexing and diversity coding.

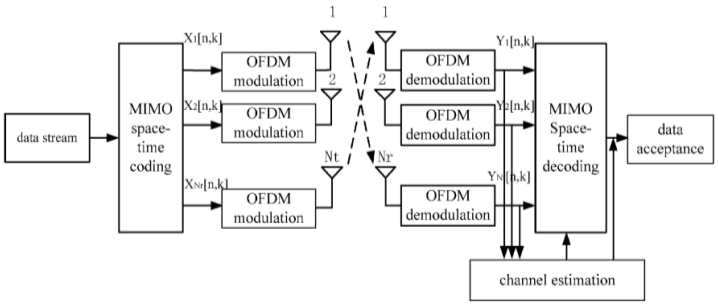


Fig. 1 Block diagram of MIMO-OFDM System

(B) Channel Estimation in Single Input Single Output (SISO) System

This analysis focuses on a single input, single output (SISO) system. Channels can affect different frequency signals in unique ways, making it essential to estimate the channel for each frequency channel. Due to the potentially large number of channels, this task can become complex and resource-intensive. Therefore, channel estimation is often performed on a subset of channels, with the estimates for the remaining channels being interpolated from these calculated values.

Consider selecting three frequencies, f_1, f_2 and f_3 from all the available frequencies. The corresponding reference or pilot signals will be $x(f_1)$, $x(f_2)$, and $x(f_3)$. These reference signals are complex, consisting of I/Q data. For the other frequency channels, we can interpolate the channel characteristics using the calculated ones. When these reference signals are received at the destination, they will include distortions and noise, represented by $y(f_1)$, $y(f_2)$, and $y(f_3)$. To express the received signals y(f) in terms of x(f), a channel function h(f) is required for the specific frequency. The relationship between the reference signal, the received signal, and the channel function can thus be represented by the correlation function.

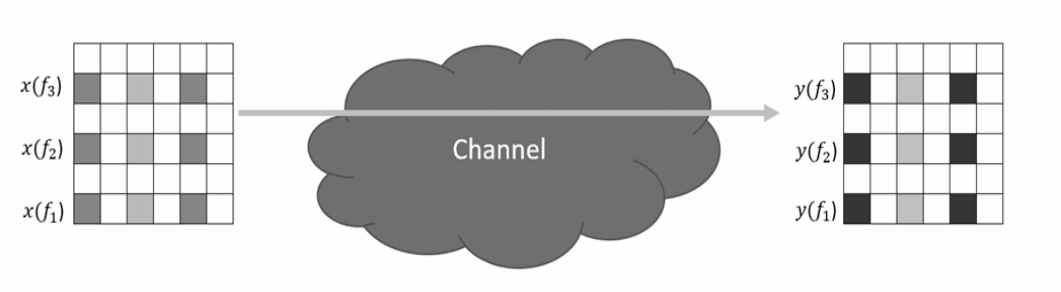


Fig. 2 Single Input Single Output System

Which can be written as

$y(f_1) = h(f_1) \cdot x(f_1)$
$y(f_2) = h(f_2) \cdot x(f_2)$
$y(f_3) = h(f_3) \cdot x(f_3)$
….(1)

from these equations, x(f) and y(f) are known thus h(f) can be calculated

$h(f_1) = y(f_1) \cdot x^H(f_1)$
$h(f_2) = y(f_2) \cdot x^H(f_2)$
$h(f_3) = y(f_3) \cdot x^H(f_3)$
….(2)

Here $x^H(f)$ is the Hermitian of $x(f)$.

Since only three frequencies have been considered and channel characteristics have been estimated for these frequencies, the properties for the remaining frequencies can be interpolated from the known characteristics. The channel function in these equations represents channel distortion. Additionally, noise is added to the distorted signals, so the actual equations of the received signal are as follows:

$y(f_1) = h(f_1) \cdot x(f_1) + n(f_1)$
$y(f_2) = h(f_2) \cdot x(f_2) + n(f_2)$
$y(f_3) = h(f_3) \cdot x(f_3) + n(f_3)$ ….(3)

Like how channel function was estimated, theoretically noise can also be estimated by using averaged channel estimate $\hbar(f)$.

$n(f_1) = y(f_1) - \hbar(f_1) \cdot x(f_1)$
$n(f_2) = y(f_2) - \hbar(f_2) \cdot x(f_2)$
$n(f_3) = y(f_3) - \hbar(f_3) \cdot x(f_3)$ ….(4)

However, absolute values of noise are not helpful for channel estimation due to the continuous variations in noise within a channel. What is beneficial is an estimated function of noise that accounts for and models these variations. Various algorithms and methods can achieve this. One such method, implemented in srsLTE (an open-source LTE implementation on SDR), involves subtracting the averaged channel estimate from the actual channel estimate.

$n(f_1) = h(f_1) - \hbar(f_1)$
$n(f_2) = h(f_2) - \hbar(f_2)$
$n(f_3) = h(f_3) - \hbar(f_3)$
….(5)

(C) Channel Estimation in Multiple Input Multiple Output (MIMO) System

In a multiple input multiple output (MIMO) system, there are multiple transmission endpoints and multiple reception endpoints.

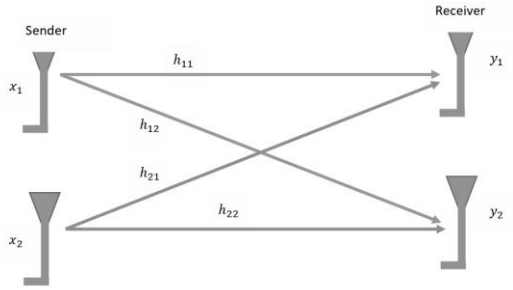


Fig. 3 A general 2x2 MIMO System

In a MIMO system, the channel estimation process is similar, but now there are two signals received from a single source. This indicates that two distinct paths in the medium were used, one for each signal. Therefore, to compute the final signal y(f) for each frequency, both received signals must be considered. This results in the formation of a matrix of received signals.

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

….(6)

Similar to a SISO system, in a MIMO system, the Hermitian of the input matrix x can be taken to estimate the channel matrix h. Similarly, the noise matrix n can also be calculated using matrix operations.

3. Channel Estimation

(A) Least Square (LS) Channel Estimation

Least squares (LS) channel estimation is a fundamental technique used in wireless communication systems to estimate channel state information (CSI). It is one of the simplest and most widely used methods due to its straightforward implementation and solid theoretical foundation. The LS channel estimation method aims to minimize the sum of the squared differences between the observed received signals and the predicted signals based on the estimated channel. This approach leads to a straightforward and intuitive solution, making it easy to implement.

LS channel estimation for 5G multicarrier wireless communications in Rayleigh, Rician and Nakagami fading channel is given by

$\hat{H}_{LS} = (X^H X)^{-1} X^H Y = X^{-1} Y$
….(7)

Where estimate of H is $\hat{H}$ and X^H is the Hermitian matrix of X.

(B) Minimum Mean Square Error (MMSE) Channel Estimation

Minimum Mean Square Error (MMSE) channel estimation is an advanced technique used in wireless communication systems to estimate channel state information (CSI) with improved accuracy and robustness compared to simpler methods like Least Squares (LS) estimation. MMSE channel estimation aims to minimize the mean square error between the actual and estimated channel responses by incorporating statistical knowledge

of the channel and noise. This method leverages both the received signal and prior information about the channel statistics to achieve a more accurate estimate.
MMSE channel estimation, the received signal y is modelled as
y = Hx + n,
where H is the channel matrix to be estimated, x is the transmitted signal, and n is the noise. The MMSE estimate $\hat{H}$ is given by:
$\hat{H}_{MMSE} = R_H X^T (X R_H X^T + \sigma^2 I)^{-1} y$
….(8)
Where R_H is the autocorrelation matrix of the channel, σ^2 is the noise variance, and I is the identity matrix.

(C) M-Estimator Channel Estimation

An M-estimator [16] is employed to enhance the robustness of channel estimation techniques in 5G wireless communication systems. The penalty, influence, and weight functions of the M-estimator for channel estimation in 5G multicarrier wireless communication over Rayleigh and Rician fading channels are provided by

$$\rho_{M-estimator}(x) = \begin{cases} \frac{x^2}{2} \ for\ |x| \le a \\ a^2 - a|x| \ for\ a < |x| \le b \\ \frac{-ab}{2} x\, exp\left(1 - \frac{x^2}{b^2}\right) + d \ for\ |x| > b \end{cases}$$

….(9)

$$\psi_{M-estimator}(x) = \begin{cases} x \ for\ |x| \le a \\ a\, sign(x)\ for\ a < |x| \le b \\ \frac{a}{b} x\, exp\left(1 - \frac{x^2}{b^2}\right) \ for\ |x| > b \end{cases}$$

….(10)

$$w_{M-estimator}(x) = \begin{cases} 1 \ for\ |x| \le a \\ \frac{a\, sgn(x)}{x} \ for\ a < |x| \le b \\ \frac{a}{b} exp\left(1 - \frac{x^2}{b^2}\right) \ for\ |x| > b \end{cases}$$

….(11)
Where d is a constant. The choice of the constants a(=kv^2) and b(=$2kv^2$) depends on the robustness measures derived from the influence function (k is also any constant). A robust estimator should possess a finite value of c. M- estimators reduce the outcome of outliers with $min \sum_i \rho(x_i)$, where ρ(x) is penalty function, influence function ψ(x) is the derivative of penalty function and $w_{Mestimator}(x)=\frac{\psi(x)}{x}$ is the weight function. The penalty function, influence function and weight function of M-estimator are shown in figures 4 – 6

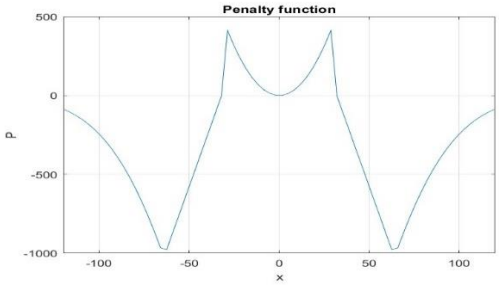


Fig. 4 Penalty Function of M-estimator

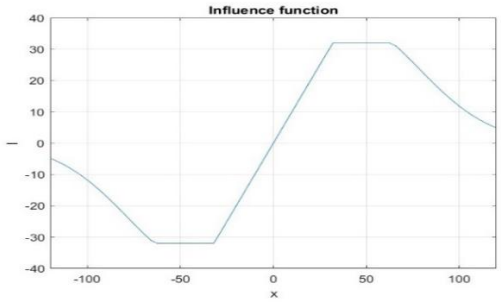


Fig. 5 Influence Function of M-estimator

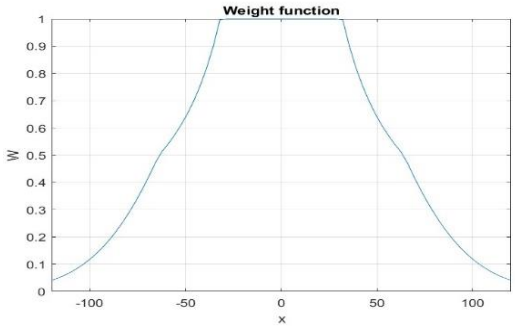


Fig. 6 Weight Function of M-estimator

M-estimator channel estimation for 5G multicarrier wireless communications in Rayleigh, Rician and Nakagami fading channel is given by

$$\hat{H}_{M-Estimator} = (X^H \psi(x) X)^{-1} X^H Y$$

….(12)

Where ψ(x) is influence function of the estimator described in above section. X is the transmitted signal in the channel. H is Hermitian transform matrix. Y represents the received signal in the channel.

(D) Modified M-estimator-I

From the statistics we know that the M-estimator is a statistical estimator used in robust statistics to estimate parameters of a statistical model. It is particularly useful when the data may contain outliers or other departures form the assumed model. In the context of wireless channel estimation, the M-estimator approach can be applied to estimate parameters such as the channel impulse response (CIR), the channel frequency response (CFR), or the channel matrix in MIMO systems. The previously proposed M-estimator based channel estimation algorithm uses the influence function of M-estimator. The modification proposed in M-estimator, which is named as Modified M-estimator-I uses ratio function instead of influence function.

The ratio function is described as the ratio of influence function to penalty function or object function. The ratio function is obtained as $R_{MODIFIED}(x) = \frac{\psi_{PROPOSED}(x)}{\rho_{PROPOSED}(x)}$. The proposed ratio function equation is described below

$$R_{MODIFIED}(x) = \begin{cases} \frac{2}{x} & for\ |x| \le a \\ \frac{sign(x)}{a-|x|} & for\ a < |x| \le b \\ \frac{\frac{a}{b}x\ exp\left(1-\frac{x^2}{b^2}\right)}{\frac{-ab}{2}x\ exp\left(1-\frac{x^2}{b^2}\right)+d} & for\ |x| > b \end{cases}$$

….(13)

The Modified M-estimator 2 based method finds the channels estimation parameters using the below equation

$$\hat{H}_{MODIFIED-I} = (X^H R(x) X)^{-1} X^H Y$$

….(14)

Where The R(x) is the ratio function described in above section. Ratio function is piecewise function, where the expression changes depending on the interval in which x lies. The X is the transmitted signal in the channel. H is Hermitian transpose matrix. Y represents the received signal in the channel. Figure… shows the ratio function of proposed Modified M-estimator-I.

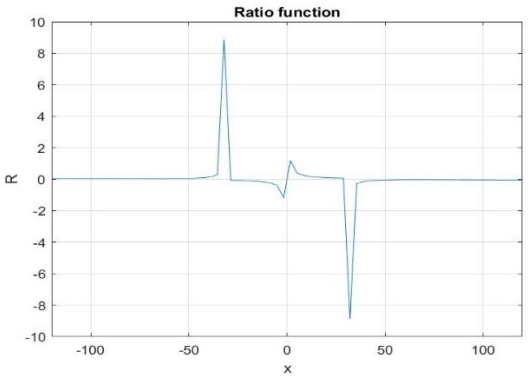


Fig. 7 Ratio Function of Proposed Modified M-estimator-I

4. Simulation Results

In simulations, channel estimation techniques based on LS (Least Squares), MMSE (Minimum Mean Square Error), M-estimator, and Modified M-estimator-I are compared. The performance evaluation of these techniques is conducted in terms of Mean Squared Error (MSE) and Signal-to-Noise Ratio (SNR). Figures 8 to 10 present the MSE and SNR results of these estimation techniques in Rayleigh fading, Rician fading, and Nakagami fading channels, respectively. It is evident across all these algorithms that MSE decreases as SNR increases, indicating better performance at higher SNR levels. Simulation results demonstrate that the Modified M-estimator-I outperforms LS, MMSE, and M-estimator techniques in channel estimation. All coding and simulations for channel estimation are implemented in MATLAB. The simulation parameters are detailed in the table below.

Table-1 Details of Simulation Parameters

Sr. No.	Parameters	Value(s)
1	Number of base station antennas	2
2	Number of UE antennas	2
3	Number of the resource block	6
4	Number of the resource bock per signal carrier	12
5	Number of transmitted symbols	2000
6	DFT size	Number of resource block X number of resource blocks per single carrier
7	FFT size	500
8	OFDM CP	Normal
9	Power delay profile	[1, 0.5, 0.25, 0.125]
10	Channel	Rayleigh fading channel, Rician fading channel, Nakagami fading channel
11	Channel estimation algorithms	LS, MMSE, M-estimator, Modified M-estimator

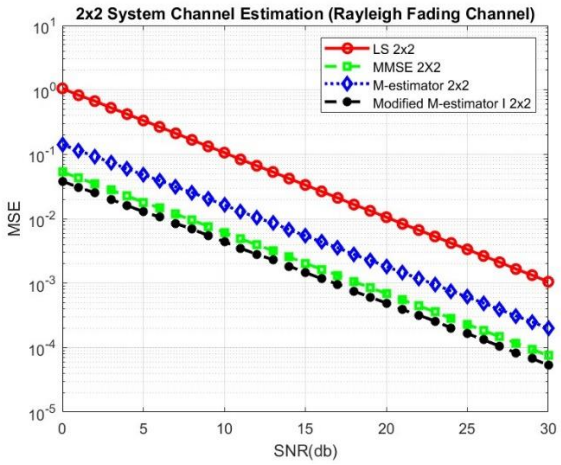


Fig. 8 BER vs SNR Performance of LS, MMSE, M-estimator and Proposed modified M-estimator-I in Rayleigh Fading channel.

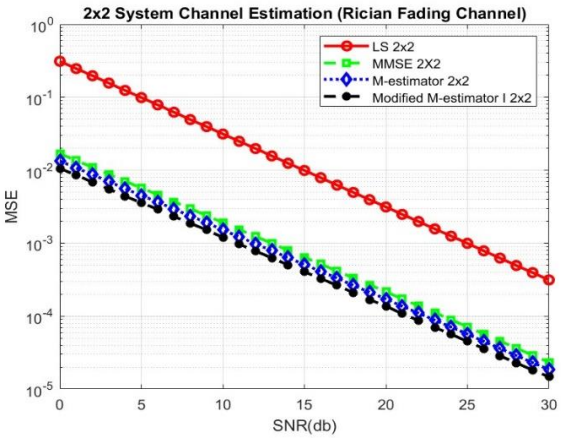


Fig. 9 BER vs SNR Performance of LS, MMSE, M-estimator and Proposed modified M-estimator-I in Rician Fading channel.

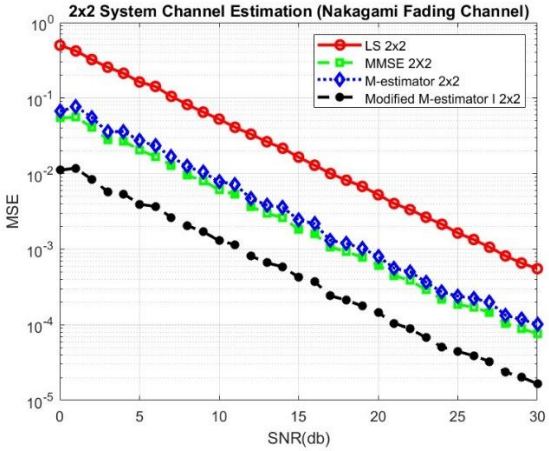


Fig. 10 BER vs SNR Performance of LS, MMSE, M-estimator and Proposed modified M-estimator I in Nakagami Fading channel.

5. Conclusion

In this paper, analysis of a Modified M-estimator-I based channel estimation technique for OFDM wireless communication systems in Rayleigh, Rician, and Nakagami fading channels is done. The newly proposed Modified M-estimator-I is employed to optimize the channel estimation process. To validate its effectiveness, we

provide simulation results that compare the performance of our proposed technique against the LS, MMSE, and M-estimator techniques. Based on these simulations, it is observed that the proposed Modified M-estimator-I based channel estimation technique demonstrates superior performance compared to LS, MMSE, and M-estimator techniques across Rayleigh, Rician, and Nakagami fading channels.

6. References

1. Malik, P.K., Wadhwa, D.S. & Khinda, J.S., "A Survey of Device to Device and Cooperative Communication for the Future Cellular Networks". Int J Wireless Inf Networks (2020). https://DOI.org/10.1007/s10776-020-004828
2. M. Morelli and U. Mengali, "A comparison of pilot-aided channel estimation methods for OFDM systems," in IEEE Transactions on Signal Processing, vol. 49, no. 12, pp. 3065-3073, Dec. 2001, DOI: 10.1109/78.969514.
3. Imran Khan, Joel J. P. C. Rodrigues, Jalal Al-Muhtadi, Muhammad Irfan Khattak, Yousaf Khan, Farhan Altaf, Seyed Sajad Mirjavadi, Bong Jun Choi, "A Robust Channel Estimation Scheme for 5G Massive MIMO Systems", Hindawi, Wireless Communications and Mobile Computing Volume 2019, Article ID 3469413.
4. Y. Li, L. J. Cimini and N. R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," in IEEE Transactions on Communications, vol. 46, no. 7, pp. 902-915, July1998.DOI: 10.1109/ 26.701317
5. C. Yin, J. Li, X. Hou and G. Yue, "Pilot Aided LS Channel Estimation in MIMO-OFDM Systems," 2006 8th international Conference on Signal Processing, Guilin, China, 2006, DOI: 10.1109/ICOSP.2006.345832
6. T. A. Kumar and L. Anjaneyulu, "Channel Estimation Techniques for Multicarrier OFDM 5G Wireless Communication Systems," 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 2020, pp. 98-101, DOI: 10.1109/ICSET51301.2020.9265353.
7. Ye Li, "Pilot-symbol-aided channel estimation for OFDM in wireless systems," in IEEE Transactions on Vehicular Technology, vol. 49, no. 4, pp. 1207-1215, July 2000. DOI: 10.1109/25.875230
8. Motade, S.N.; Kulkarni, A.V. Channel Estimation and Data Detection Using Machine Learning for MIMO 5G Communication Systems in Fading Channel. Technologies 2018, *6*, 72. https://DOI.org/10.3390/technologies 6030072
9. M. Jalloh, M. Al-Gharabally and P. Das, "Performance of OFDM Systems in Rayleigh Fading Channels with Phase Noise and Channel Estimation Errors," MILCOM 2006 - 2006 IEEE Military Communications conference, Washington, DC, 2006, pp.1-7. DOI: 10.1109/MILCOM .2006.302360
10. T. Anil Kumar and K. Deergha Rao, "A New M–Estimator Based Robust Multiuser Detection in Flat-Fading Non-Gaussian Channels," IEEE Trans. On Comm., pp. 1908- 1913, Vol. 57, No. 7, July 2009.
11. V. Durga and S. Anuradha, "On Channel Estimation in Universal Filtered Multi-Carrier (UFMC) System," 2019 PhotonIcs & Electromagnetics Research Symposium - Spring (PIERS-Spring), Rome, Italy, 2019, pp. 3708-3713, DOI: 10.1109/PIERS-Spring46901.2019.9017701.
12. Navid daryasafar; Aboozar lashkari; Babak ehyaee. Channel estimation in MIMO-OFDM systems based on comparative methods by LMS algorithm. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 3, No 3, May 2012
13. Qin, Qibo et al. "Sparse Channel Estimation for Massive MIMOOFDM Systems Over Time-Varying Channels." IEEE Access 6 (2018): 33740-33751.
14. C. Chen and W. Wu, "Joint AoD, AoA, and Channel Estimation for MIMO-OFDM Systems," in IEEE Transactions on Vehicular Technology, vol. 67, no. 7, pp. 5806-5820, July 2018. DOI: 10.1109/TVT.2018. 2798360
15. T. Anil Kumar and L. Anjaneyulu, "Implementation of Best Hybrid Adaptive and Intelligent MIMO Detector on Reconfigurable Architecture for 5G LTE/IoT Environment," International Conference on Smart Trends for Computing and Communications (SmartCom 2020), 24th - 25th January 2020, PP 49-56, Bangkok, Thailand. 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET), 9
16. T. A. Kumar and L. Anjaneyulu, "Channel Estimation Techniques for Multicarrier OFDM 5G Wireless Communication Systems," 2020 IEEE 10th International Conference on System Engineering and Technology (ICSET), Shah Alam, Malaysia, 2020, pp. 98-101, DOI: 10.1109/ICSET51301.2020.9265353.

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 02 2024

Integration of deep learning and plant biosecurity toward sustainable agriculture: A SWOT analysis

Abhilasha Sharma; Parul Sharma

Check for updates

AIP Conf. Proc. 3168, 020011 (2024)
https://doi.org/10.1063/5.0219003

View Online
Export Citation

05 July 2024 06:08:08


APL Energy
Latest Articles Online!
Read Now
AIP Publishing

AIP Publishing

Integration of Deep Learning and Plant Biosecurity Toward Sustainable Agriculture: A SWOT Analysis

Abhilasha Sharma and Parul Sharma [a)]

Software Engineering Department, Delhi Technological University, Delhi, India

[a)] Corresponding author: sharma05oct@gmail.com

Abstract. Plant biosecurity holds immense significance for every nation as it plays a crucial role in protecting crops, ensuring food security, and safeguarding ecology and the livelihoods of individuals. In this regard, plant biosecurity constitutes a vital component of sustainable agriculture progress. Digitalization is the new go-to strategy to address agriculture's productivity, sustainability, and resilience challenges. Deep Learning and other digital technologies that have emerged in recent years are currently being applied in every agriculture practice. Researchers from multidisciplinary areas strive to use these technologies and propose novel solutions to expedite the in-field workflow. However, struggle to put their solutions into production, deliver tangible results and obtain favorable outcomes with limited in-field datasets. In this view, this paper reviews remarkable research integrating deep learning with plant biosecurity and provides a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis to identify the strengths, weaknesses, opportunities, and threats inherent to deep learning-driven solutions as facilitators or barriers to plant biosecurity.

INTRODUCTION

Agriculture holds great importance in human civilization and is the cornerstone of modern society. Its profound connections to the global economy, biodiversity, and human history are undeniable [1]. As the most significant industry worldwide, it employs over 1.3 billion individuals and is valued at $5 trillion [2]. Agriculture enabled humans to cultivate their food, thus propelling the transition from a hunting society to a self-sustaining one. With the advent of digital technologies and the industrial revolution, the agriculture era has changed its paradigm from conventional to sustainable agriculture [3]. The agricultural transformation is represented in Fig. 1.

05 July 2024 06:08:08

Plant biosecurity is a critical use case of sustainable agriculture [4]. The pressures of population growth, environmental destruction, declining food security index, and climate change adversely impact agricultural sustainability [5]. Extreme climate change continues to raise the plant host's susceptibility to pests, diseases, and weeds, negatively affecting crop yields and quality. Pests, diseases, and weeds account for a massive loss in crop productivity and their nutritious values. Farmers invest a lot of money and time in preventing the abovementioned abnormalities; however, symptoms are hard to perceive through human sight, which may lead to total crop failure. Though conventional agriculture techniques provide some solutions, including applying chemicals and manual removal of weeds, these techniques are subjective to farmers' knowledge, time-consuming, and have adverse environmental effects. [6]. To this extent, smart agriculture, utilizing the techniques of Deep Learning (DL), has evolved to automate agricultural practices. DL algorithms empower plant biosecurity by automating diagnosis and decision-making [7-9].

However, many attempts at existing DL-based models and frameworks for plant biosecurity fail in natural field conditions. Despite the tremendous DL-based agricultural transformation in the last decade, NASSCOM estimates that only 2% of Indian farmers use technological solutions. Therefore, this paper aims to bridge the gap between recent developments and practices of DL in plant biosecurity. A comprehensive SWOT analysis, known as the Strengths, Weaknesses, Opportunities, and Threats, is conducted to accomplish this goal; it is worth mentioning here that this work is the first attempt to showcase a SWOT of this domain. This assessment examines the internal and external factors that influence the growth and potential of the area, encompassing both positive and negative aspects. The paper is further structured as follows: the next section describes the various terminologies constituting the research area. A brief overview of the SWOT matrix is presented in a further section. The consequent section explains the SWOT analysis of deep learning in plant biosecurity, followed by another section, which concludes the article.

1st International Conference on Recent Advancements in Computing Technologies & Engineering
AIP Conf. Proc. 3168, 020011-1–020011-11; https://doi.org/10.1063/5.0219003
Published under an exclusive license by AIP Publishing. 978-0-7354-4990-9/$30.00

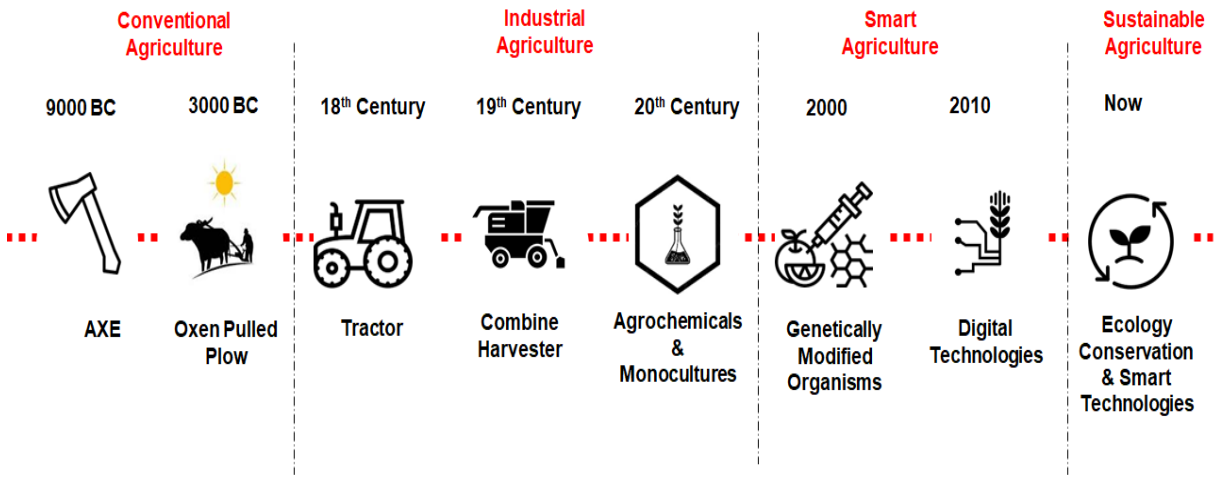


FIGURE 1. Evolution of Agriculture

TERMINOLOGIES

This section describes various terminologies that constitute the proposed research topic: Sustainable Agriculture, Plant Biosecurity, Smart Agriculture, and Deep Learning.

Sustainable Agriculture

Sustainability refers to fulfilling our current requirements without compromising future generations' needs. It enhances the quality of life, safeguards ecosystems, and preserves natural resources for the well-being of future generations. Given agriculture's profound ties to the global economy, human societies, and biodiversity, it is one of the most critical worldwide conservation domains. Hence sustainability in agricultural practices is the foremost need [10].

Sustainable agriculture involves employing environmentally friendly practices to meet the current demands for food and textiles while preserving resources for the coming generation. It encompasses farming methods that prioritize the protection of human and natural systems, ensuring the production of crops without causing harm to ecology. Sustainable agriculture rests on three essential pillars: environmental health, economic profitability, and social equity. Figure 2 presents the six primary components of sustainable agriculture: plant biosecurity, crop diversity, soil management, crop rotation, water management, and nutrient management [11].

Plant Biosecurity

Biosecurity implies implementing preventive measures to hinder the introduction and spread of harmful organisms such as viruses, bacteria, and other pathogens, aiming to minimize the risk of infectious disease transmission. The primary objective of biosecurity is to prevent, control, and manage risks associated with life and health, tailored to the specific sector of biosecurity. These measures safeguard crops against pests, invasive species, and other organisms threatening human and plant welfare. Plant biosecurity comprises a range of standards designed to shield crops from existing and exotic plant pests, diseases, and weeds at various levels, including national, regional, and individual farms [13].

Plant biosecurity is pivotal in ensuring food safety, food security, trade, market access, and development. It significantly influences the profitability and sustainability of plant industries [14]. Figure 3 illustrates the critical tasks involved in plant biosecurity, *Prevention* being the core focus to prevent the entry of exotic pests and diseases across boundaries. *Early Detection and Surveillance* indicates monitoring of pests, disease, and weeds to take prompt response against their growth; *Response and Management* imply the usage of various methods to mitigate the impacts

05 July 2024 06:08:08

and prevent further spread.; *Research and Innovation* suggest developing improved decision-making and risk analysis tools; *Collaboration and Cooperation* indicate communication and coordination among socio-economic drivers.



FIGURE 2. Elements of Sustainable Agriculture

Smart Agriculture

Smart agriculture refers to using digital technologies like the Internet of Things, sensors, location systems, robots, and artificial intelligence for automation and improving farm practices to increase crop quality & quantity while optimizing human labor. The penetration of the above-mentioned digital technologies brings many opportunities to enable a sustainable future in agriculture. These technologies play an essential role in achieving seven primary aspects of smart agriculture- digital soil analysis, intelligent irrigation, climate monitoring, agricultural robots, digital market intelligence, crop inspection, and crop protection [15,16]. The digital technologies that transformed agriculture as smart agriculture are represented in Fig 4.

05 July 2024 06:08:08

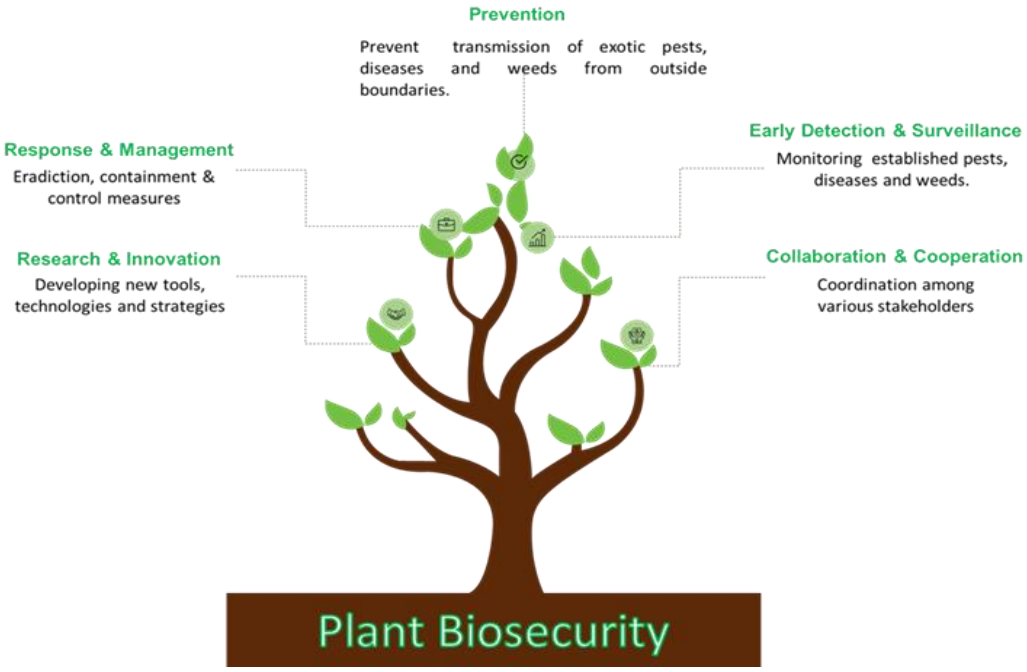


FIGURE 3. Elements of Plant Biosecurity Towards Safeguarding Plant Health



FIGURE 4. Key Technologies for Smart Agriculture

DL has recently emerged as a promising technology in intelligent agriculture, making a widespread digital impact. Deep learning refers to a class of artificial intelligence algorithms that utilize neural networks consisting of multiple layers to extract features from data and make predictions about new data. In artificial neural networks, learning involves adjusting the connection weights between neurons to improve network performance. Learning in artificial

05 July 2024 06:08:08

neural networks can be categorized into three main types: supervised, unsupervised, and reinforcement learning [17-19].

Supervised learning involves providing labeled input data to the model for training purposes and testing on unlabelled data. On the other hand, unsupervised learning focuses on identifying patterns and structures in unlabelled data, where only input data is provided to the model. Reinforcement learning enables an agent to learn in an interactive environment through trial and error, using feedback derived from its actions and experiences. Figure 5 depicts the relationship between Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL).

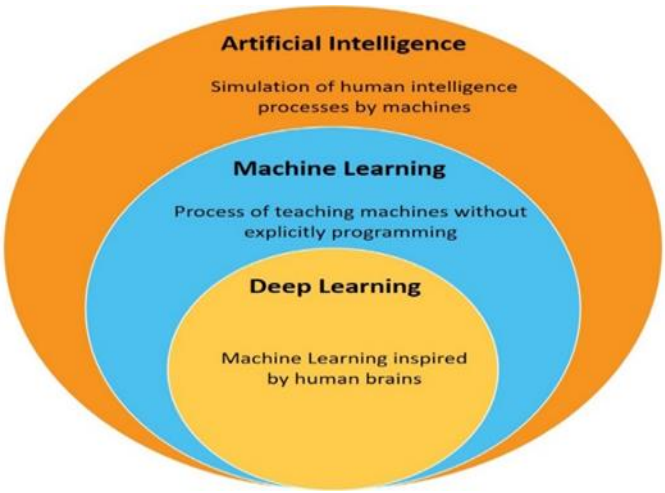


FIGURE 5. Cascading of Artificial Intelligence, Machine Learning, and Deep Learning

SWOT MATRIX- AN OVERVIEW

SWOT stands for Strengths, Weaknesses, Opportunities, and Threats. A SWOT matrix is a structured planning tool to assess an area's strengths, weaknesses, opportunities, and threats [20]. This analytical technique forms the basis for evaluating the internal capabilities and limitations, as well as the potential opportunities and threats stemming from the external environment. It considers all the positive and negative factors that impact the growth and scope of a particular field. The SWOT analysis encompasses the following scenarios: internal strengths & weaknesses and opportunities & threats in the external environment.



FIGURE 6. Generic SWOT Matrix

05 July 2024 06:08:08

SWOT ANALYSIS FOR THE INTEGRATION OF DEEP LEARNING & PLANT BIOSECURITY TO ACHIEVE AGRICULTURE SUSTAINABILITY

This section provides a structured SWOT analysis of DL-driven trends that can play an essential role in achieving plant biosecurity to strengthen agriculture sustainability. The SWOT matrix highlights the potential contributions, opportunities, weaknesses, and threats toward the applicability of DL in plant biosecurity. It provides recommendations to the researchers for an aligned use of DL for plant biosecurity. The following subsections expound on the details.

Strengths

Early Outbreak Warning

In recent years, the need for fast and accurate solutions for pests, diseases, and weeds management has been rising for better, reliable, and precise decision-making. Early-warning systems play a crucial role in plant protection, offering valuable tools that provide farmers with timely forecasts. Due to advances in computer vision and DL, the identification of crop disorders has shifted from conventional techniques to optical identification (use of digital images). DL-driven solutions provide accurate and swift detection methods and hold great potential that enables timely treatment. This, in turn, can significantly mitigate total crop loss failure, reducing financial setbacks and ensuring food security [21].

Improves Crop Yield and Lowers Production Costs

Agricultural production must rise by 70% by 2050 to meet the increasing global food demand. Managing pests, diseases, and weeds is paramount for food security. Plant disorders can significantly lower crop productivity. Traditionally, diagnosing crop abnormalities relied on human experts, which was expensive and time-consuming. Deep learning resolves the issues mentioned above and is employed to provide solutions for automated crop abnormalities detection without expert scouting. It improves productivity and lowers the cost incurred in traditional scouting techniques and blanket spraying of costly chemicals on plants [22].

Healthier Food Quality

Crop abnormalities affect crop yield and their nutritious value. The rise in plant diseases, pests, and weed outbreaks threatens food safety in various regions worldwide. Simultaneously, a global pandemic is endangering the health and well-being of millions of people on our planet. Further, this situation is worsened by farmers' blanket spraying of chemicals, pesticides, herbicides, and other chemicals without precise knowledge about these abnormalities. This has a significant impact on consumers' health and ecology. The DL-based automated solutions contribute towards maintaining nutritious food values by timely identifying phenotypic changes in plants and suggesting the exact abnormality and its possible solution [23].

Economic Benefits to Farmers, Consumers, and Country

Plant health has become a pressing global issue, impacting food crop production and nations' social and political stability. Agricultural productivity plays a vital role in the economic growth of every country. However, plant illnesses significantly affect agricultural production, resulting in substantial losses in the global economy. Monitoring plants' health and diagnosing disorders is crucial for promoting sustainable agriculture and enhancing trade. Automated DL-based early diagnosis solutions have emerged as valuable tools in disease control, aiding in selecting appropriate control techniques to improve productivity. Ultimately, this improvement benefits farmers, consumers, and the overall economy of nations [24].

Low Environmental Impacts

Agricultural practices and ecology are interconnected. Pests, diseases, and weeds are often controlled with artificial chemicals, which are harmful to producers, consumers, and ecology if not used in the correct amount. Hence to strengthen ecology, agricultural activities need to be redesigned. The introduction of DL in agricultural production is revolutionizing the farming sector. This advancement optimized inputs such as fertilizers, pesticides, herbicides, and other chemicals. By adopting DL-grounded solutions, farmers are motivated to reduce their environmental impact by adopting sustainable practices, resulting in a notable 21% reduction in greenhouse gas (GHG) emissions through the more efficient use of chemicals [26].

05 July 2024 06:08:08

Weaknesses

High Upfront Cost

DL has great potential to address the increasing food demand. However, the adoption of this technology in plant biosecurity has been relatively slow due to various factors. These include the high initial investment, limited technical expertise, and growing data privacy concerns. As DL applications tend to be computationally intensive, the requirement for substantial processing power is one of the major weaknesses. Additionally, DL-grounded solutions heavily rely on large amounts of data leading to significant costs associated with data processing. Overcoming these challenges is crucial for the widespread adoption of DL in plant health monitoring. By addressing the above-mentioned issues, plant biosecurity can fully leverage the potential of DL to meet the growing food demand effectively [26].

Low Awareness Among Farmers

Researchers are proposing automated solutions for plant abnormalities. However, low awareness among producers(farmers) about these tools and solutions is another hurdle. Easy comprehension, understanding, low access, and operational costs of radio broadcasts make them affordable for farmers seeking valuable insights regarding crop health. The automated DL solutions are being accessed by educated farmers only; less educated farmers are not exploiting these solutions because they need to gain basic knowledge about the usage of these solutions. Additionally, low-income farmers with small land areas have low interest in using these automated solutions [27].

Slow Adoption Rate in Developing Countries

Recently, DL has been addressing the management of pests, diseases, and weeds that hinder crop quality and quantity; however, only a few examples showcase positive impacts on rural livelihoods; these successes have yet to reach the expected scale. Inadequate infrastructure and resources, unreliable internet connections, and a need for more skilled professionals in the field are significant limitations. Overcoming these barriers is crucial for promoting the widespread adoption of DL in plant phenotyping. Efforts should focus on raising awareness, improving infrastructure, expanding resources, strengthening internet connectivity, and fostering the development of skilled professionals to maximize the potential benefits of DL for rural communities [27].

Non-Availability of In-Field Dataset

Most existing DL-based solutions to identify pests, diseases, and weeds are built using datasets captured in controlled conditions or considered only specific crops or diseases/pests/weeds. However, any technique used in practice must be prepared to deal with different environmental circumstances, diverse crops, pests, diseases, and weeds. Further, most self-collected datasets have yet to be released to be fully exploited for further research [7-9].

Heavy Models

DL-based solutions typically demand significant computational power and resources, making them less suitable to deploy on lightweight devices like mobiles and the Internet of Things (IoT). However, training these models can be exceedingly expensive due to the data models' complexity. Furthermore, the requirement for costly GPUs adds to the overall cost for developers and users. Additionally, efficient tuning of hyperparameters is necessary for optimizing the performance of these models [7-9].

Low Generalizability of Automated Solutions

As discussed above, most existing solutions are trained on laboratory-based datasets. Hence, their accuracy drops when these models are utilized for inference in field conditions. Further, researchers considered some specific crops or specific abnormalities, but in actual field conditions, varieties of crops and abnormalities exist. Hence a model/ solution developed for a particular crop is not generalizable to other crops.

05 July 2024 06:08:08

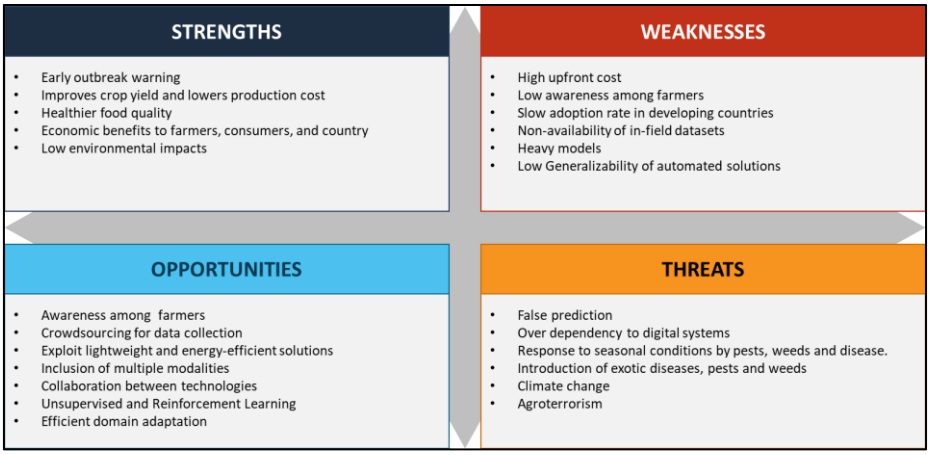


FIGURE 7. SWOT Matrix for Deep Learning in Plant Biosecurity

Opportunities

Awareness Among Farmers

To fully exploit the potential of DL for plant biosecurity, it would be beneficial if the government could circulate the various policies, websites, and apps for plant phenotyping through media. Digital literacy programs should be conducted for farmers. Farmers should also be encouraged to employ DL-based solutions to save their crops from pests, diseases, and weeds and make precise control decisions. Further, infrastructure (smartphones) should be made available to low-income farmers.

Crowdsourcing for Data Collection

A significant area for improvement in this field is the unavailability of standard datasets for DL. To make a model generalizable, it is required that researchers have datasets that are collected from diverse graphical conditions with varying weather conditions, illumination conditions, diversity of pests, diseases & weeds, and many more. Researchers alone cannot collect this data; hence crowdsourcing can generate many training data of good quality through social media or any other medium [28].

Exploit Lightweight and Energy-Efficient Solutions

As discussed above, heavy DL models cannot be deployed on embedded devices and mobile phones to be further utilized by farmers as a decision-making tool. Heavy models consume substantial energy resources that hinder the ecology facet. Considering these weaknesses, developing lightweight and energy-efficient models opens various opportunities for DL models to be fully utilized for plant biosecurity to improve agriculture sustainability [7-9].

Inclusion of Multiple Modalities

Multimodal learning is where models are trained to process and analyze multiple forms of input data, such as text, images, and audio. Most existing models primarily focus on visual information in embodiments for plant biosecurity, neglecting the potential insights from other modalities. However, incorporating multiple modalities holds immense potential for enhancing the monitoring of pests, diseases, and weeds [21-24]

Collaboration Between Technologies

To fully leverage the benefits of DL in plant biosecurity, collaboration among various technologies is essential. While DL has the potential to revolutionize this field, it is still evolving, and costs can be high. Therefore, ensuring widespread adoption will necessitate collaboration among different technologies. Integrating DL approaches with unmanned aerial vehicles (UAVs) and spectral data can facilitate the development of advanced intelligent solutions. Further, cloud computing can be integrated with DL to handle resource-intensive models, while edge computing can be combined with DL to reduce latency during inference. These are just a few examples of potential collaborations that can be explored to enhance the effectiveness and efficiency of plant biosecurity efforts [21-24].

Unsupervised and Reinforcement Learning

Existing DL solutions for plant biosecurity are based on supervised learning. However, the potential of unsupervised and reinforcement learning still needs to be addressed. To overcome the barriers of requiring large amounts of annotated data and building a heavy model for extensive labelled data can be addressed by exploring the solutions based on unsupervised and reinforcement learning [21-24].

Efficient Domain Adaptation

Domain adaptation is a computer vision field that aims to train a neural network on a source dataset and achieve high accuracy on a target dataset that significantly differs from the source dataset. The purpose of domain adaptation is to

05 July 2024 06:08:08

enable a pre-trained model to perform optimally on new data without retraining on a different dataset. Rather than starting from scratch with each new dataset, domain adaptation allows the adaptation of existing models to the target dataset, thereby saving time and resources. This approach facilitates the transfer of knowledge and learning from the source to the target dataset, enabling the model to perform well even in significantly different data domains. By leveraging domain adaptation, advancements can be made in strengthening plant biosecurity measures and addressing specific challenges related to detecting and preventing pests, diseases, and other threats to plant health [21-24].

Threats

False Prediction

A model trained on a poorly annotated dataset may lead to inaccurate predictions when tested in natural field conditions. However, farmers may adopt control measures per these false predictions, which may lead to complete crop loss failure and challenges to food security. Further, the inputs (pesticides, herbicides, fertilizers, and other chemicals) utilized as control measures will have poor impacts on ecology and causes economic loss.

Over-Dependency to Digital Systems

DL-based solutions can be employed with the help of digital devices only, like computers & smartphones; this increases farmers' dependency on these digital devices to make decisions. However, these digital gadgets may lead to GHG emissions, and the model deployed may cause high energy requirements.

Response to Weather Conditions by Pests, Weeds, And Diseases

A pest, disease, or weed may respond differently in different weather and geographic conditions. Considering all weather and geographic locations in a single dataset is almost impossible. This may cause a prediction bias and lead to poor results [29].

Introduction of Exotic Pests, Diseases, and Weeds

Exotic organisms refer to organisms introduced into an area beyond their natural range and become pests in the new environment. They are also known as alien, non-native, or introduced organisms. The entry of such exotic organisms poses a significant threat to plant security, as existing models may fail to recognize them, leading to wrong forecasts. This can result in financial losses, environmental damage, food production disruptions, and adverse effects on agriculture industries [30].

Climate Change

Climate change has complex and global impacts on agricultural ecosystems. It creates favorable conditions for the proliferation of pests and diseases, particularly in temperate zones. The effects of climate change have led to economically essential crop pests becoming more destructive, posing an increasing threat to food security and the environment. With the anticipated drastic changes in climate in the future, this situation may worsen [29].

Agroterrorism

Agroterrorism involves deliberate attacks on the food supply chain, targeting livestock and crops during various stages, including production, harvesting, storage, or transport. Agroterrorism poses a significant threat to achieving plant biosecurity and ultimately challenges the sustainability of agriculture. Terrorist actions to disrupt the food supply can have far-reaching consequences on economic stability and public health. These factors highlight the critical importance of addressing measures to counter agroterrorism to safeguard plant biosecurity and ensure the sustainability of agriculture [31].

CONCLUSION

The dependency of plant biosecurity over deep learning has increased remarkably in the last decade. Deep learning has automated pests, diseases, and weeds diagnosis & helps in decision-making for farmers and plant pathologists. Despite accelerated growth in this field, the agriculture industry still needs to catch up to the slow adoption of proposed solutions. In this view, to explore this area, a SWOT analysis is performed. The SWOT analysis of deep learning in plant biosecurity has identified several strengths, weaknesses, opportunities, and threats. Based on the SWOT analysis findings, the authors recommend that the researchers must focus on its strengths and opportunities to utilize the potential of DL better and take appropriate solutions to address its weaknesses and threats. Further, the government should run digital literacy programs, provide the necessary infrastructure to low-income farmers, and circulate various policies, websites, and apps for plant phenotyping. More importantly, a collaboration among DL and other digital technologies should be established; by doing so, deep learning can further strengthen plant biosecurity to achieve agriculture sustainability.

05 July 2024 06:08:08

REFERENCES

1. L. Carlisle, M. Montenegro de Wit, M.S. DeLonge, A. Iles, A. Calo, C. Getz, J. Ory, K. Munden-Dixon, R. Galt, B. Melone, R. Knox, and D. Press, "Transitioning to sustainable agriculture requires growing and sustaining an ecologically skilled workforce", *Frontiers in Sustainable Food Systems* 3, (2019). https://doi.org/10.3389/fsufs.2019.00096
2. L. Goedde, M. Horii, and S. Sanghvi, "Pursuing the global opportunity in food and Agribusiness", *McKinsey & Company* (2015). https://www.mckinsey.com/industries/chemicals/our-insights/pursuing-the-global-opportunity-in-food-and-agribusiness
3. Y. Liu, X. Ma, L. Shu, G.P. Hancke, and A.M. Abu-Mahfouz, "From industry 4.0 to agriculture 4.0: Current status, Enabling Technologies, and research challenges", *IEEE Transactions on Industrial Informatics* 17, 4322 (2021). doi:10.1109/tii.2020.3003910
4. S.L. Vincenza Ilardi, "Plant biosecurity: Diagnosis in plant pathology" *Biosafety* 03, (2014). 10.4172/2167-0331.1000e143
5. O. Calicioglu, A. Flammini, S. Bracco, L. Bellù, and R. Sims," The future challenges of food and agriculture: An integrated analysis of trends and solutions", *Sustainability* 11, 222 (2019). https://doi.org/10.3390/su11010222
6. "New standards to curb the global spread of plant pests and diseases", *FAO* (n.d.). https://www.fao.org/news/story/en/item/1187738/icode/.
7. A. Abade, P.A. Ferreira, and F. de Barros Vidal, " Plant diseases recognition on images using convolutional neural networks: A systematic review ", *Computers and Electronics in Agriculture* 185, 106125 (2021). https://doi.org/10.48550/arXiv.2009.04365
8. N.Y. Murad, T. Mahmood, A.R. Forkan, A. Morshed, P.P. Jayaraman, and M.S. Siddiqui," Weed detection using Deep learning: A systematic literature review", *Sensors* 23, 3670 (2023). https://doi.org/10.3390/s23073670
9. A.C. Teixeira, J. Ribeiro, R. Morais, J.J. Sousa, and A. Cunha," A systematic review on automatic insect detection using Deep Learning", Agriculture 13, 713 (2023). https://doi.org/10.3390/agriculture13030713
10. A. Wezel, M. Casagrande, F. Celette, J.-F. Vian, A. Ferrer, and J. Peigné," Agroecological practices for sustainable agriculture. A Review", *Agronomy for Sustainable Development* 34, 1 (2013). doi:10.1007/s13593-013-0180-7
11. J.E. Ikerd, " The need for a system approach to sustainable agriculture ", *Agriculture, Ecosystems & Environment* 46, 147 (1993). 10.1016/0167-8809(93)90020-p
12. 1V. Renault, M.-F. Humblet, and C. Saegerman, "Biosecurity concept: Origins, evolution and perspectives", *Animals* 12, 63 (2021). doi:10.3390/ani12010063
13. R.D. Magarey, M. Colunga-Garcia, and D.A. Fieselmann, "Plant biosecurity in the United States: Roles, responsibilities, and Information Needs", *BioScience* 59, 875 (2009).
14. A.W. Sheppard, I. Gillespie, M. Hirsch, and C. Begley," Biosecurity and sustainability within the growing global bioeconomy", *Current Opinion in Environmental Sustainability* 3, 4 (2011). 10.1016/j.cosust.2010.12.011
15. M.J. O'Grady and G.M.P. O'Hare," Modelling the smart farm. Information Processing in Agriculture", *Information Processing in Agriculture* 4, 179 (2017).
16. X. Yang, L. Shu, J. Chen, M.A. Ferrag, J. Wu, E. Nurellari, and K. Huang, " A survey on Smart Agriculture: Development Modes, technologies, and security and privacy challenges", *IEEE/CAA Journal of Automatica Sinica* 8, 273 (2021). doi:10.1109/jas.2020.1003536
17. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M.S. Lew, " Deep Learning for Visual Understanding: A Review", *Neurocomputing* 187, 27 (2016). 10.1016/j.neucom.2015.09.116
18. L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep Learning for Generic Object Detection: A Survey", *International Journal of Computer Vision* 128, 261 (2019). 10.1007/s11263-019-01247-4
19. L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of Deep Learning: Concepts, CNN Architectures, challenges, applications, Future Directions ", *Journal of Big Data 8*, (2021). 10.1186/s40537-021-00444-8
20. "SWOT analysis", *Wikipedia* (2023). https://en.wikipedia.org/wiki/SWOT_analysis
21. P.S. Thakur, P. Khanna, T. Sheorey, and A. Ojha, " Trends in vision-based machine learning techniques for Plant Disease Identification: A systematic review", *Expert Systems with Applications* 208, 118117 (2022). https://doi.org/10.1016/j.eswa.2022.118117

05 July 2024 06:08:08

22. W. Li, T. Zheng, Z. Yang, M. Li, C. Sun, and X. Yang, " Classification and detection of insects from field images using deep learning for Smart Pest Management: A systematic review", *Ecological Informatics* 66, 101460 (2021). https://doi.org/10.1016/j.ecoinf.2021.101460
23. N. Rai, Y. Zhang, B.G. Ram, L. Schumacher, R.K. Yellavajjala, S. Bajwa, and X. Sun," Applications of deep learning in Precision Weed Management: A Review", *Computers and Electronics in Agriculture* 206, 107698 (2023). https://doi.org/10.1016/j.compag.2023.107698
24. Y. Zhang, M. Wang, D. Zhao, C. Liu, and Z. Liu," Early weed identification based on Deep Learning: A Review", Smart *Agricultural Technology* 3, 100123 (2023). https://doi.org/10.1016/j.atech.2022.100123
25. A. Lamba, P. Cassey, R.R. Segaran, and L.P. Koh, " Deep Learning for Environmental Conservation", *Current Biology* 29, (2019). https://doi.org/10.1016/j.cub.2019.08.016
26. V.G. Dhanya, A. Subeesh, N.L. Kushwaha, D.K. Vishwakarma, T. Nagesh Kumar, G. Ritika, and A.N. Singh, " Deep Learning Based Computer Vision Approaches for Smart Agricultural Applications. Artificial Intelligence in Agriculture", *Artificial Intelligence in Agriculture* 6, 211 (2022). https://doi.org/10.1016/j.aiia.2022.09.007
27. "The future of food and agriculture – Trends and challenges | Global Perspectives Studies | Food and Agriculture Organization of the United Nations", *Food and Agriculture Organization of the United Nations* (n.d.). https://www.fao.org/global-perspectivesstudies/resources/detail/en/c/458158/.
28. J.G. Barbedo, " Deep learning applied to plant pathology: The problem of data representativeness", *Tropical Plant Pathology* 47, 85 (2021).
29. B.K. Singh, M. Delgado-Baquerizo, E. Egidi, E. Guirado, J.E. Leach, H. Liu, and P. Trivedi, "Climate change impacts on plant pathogens, food security and paths forward", *Nature Reviews Microbiology* (2023).
30. R.D. Noar, C.J. Jahant-Miller, S. Emerine, and R. Hallberg, " Early warning systems as a component of integrated pest management to prevent the introduction of exotic pests", *Journal of Integrated Pest Management* 12, (2021).
31. R. Casagrande, "Agricultural bioterrorism (agricultural biosecurity, agroterrorism)", *Encyclopedia of Bioterrorism Defense* (2005).

05 July 2024 06:08:08

Physica Scripta

INCLUSIVE PUBLISHING
TRUSTED SCIENCE

PAPER

Investigating the effect of scaling and temperature on the performance of improved junctionless nanowire FET through simulation analysis

To cite this article: Bharti and Poornima Mittal 2024 *Phys. Scr.* **99** 086103

View the article online for updates and enhancements.

You may also like

- Oppositely-Doped Core-Shell Junctionless Nanowire FET: Design and Investigation
Bharti and Poornima Mittal
- Performance analysis of nanowire and nanosheet NCFETs for future technology nodes
Fahimul Islam Sakib, Md Azizul Hasan and Mainul Hossain
- Device design guideline for junctionless gate-all-around nanowire negative-capacitance FET with HfO_2-based ferroelectric gate stack
Yejoo Choi, Yuri Hong and Changhwan Shin

This content was downloaded from IP address 14.139.251.98 on 06/08/2024 at 07:06

Physica Scripta

PAPER

Investigating the effect of scaling and temperature on the performance of improved junctionless nanowire FET through simulation analysis

RECEIVED
31 December 2023

REVISED
3 May 2024

ACCEPTED FOR PUBLICATION
16 July 2024

PUBLISHED
29 July 2024

Bharti* **and Poornima Mittal***

Department of Electronics and Communication Engineering, Delhi Technological University, Delhi-110042, India
* Authors to whom any correspondence should be addressed.

E-mail: bhartigndit@gmail.com **and** poornima2822@gmail.com

Keywords: band-to-band tunneling (BTBT), core-shell (CS), dual material gate (DMG), high-k spacer, junctionless nanowire FET (JL-NWFET)

Abstract

An Improved Junctionless Nanowire Field Effect Transistor (I-JL-NWFET) device is proposed in this paper to address the limitations of conventional JL-NWFET. This research paper initially, comprehensively analyzes the impact of channel length (L) and channel thickness (t_{si}) scaling on the electrical, analog/RF, and linearity performance of I-JL-NWFET and JL-NWFET. The results suggest that the specific design features in I-JL-NWFET contribute to a more robust and less sensitive response to variations in scaling compared to its counterpart, JL-NWFET. Furthermore, an exploration into the impact of temperature on the electrical, analog/RF, and linearity performance is also conducted for both I-JL-NWFET and JL-NWFET. The electrical performance of I-JL-NWFET showcases a significantly reduced temperature sensitivity in parameters like drain current (I_D), Subthreshold Slope (SS) and Drain Induced Barrier Lowering (DIBL) compared to JL-NWFET. Subsequently, analyzing the analog/RF performance in the context of parameters such as transconductance (g_m), Transconductance Gain Factor (TGF), output conductance (g_d), early voltage (V_{EA}), total gate capacitance (C_{GG}), and cut-off frequency (f_T) under temperature variation, a lower degree of variability in I-JL-NWFET is observed compared to JL-NWFET. Furthermore, the linearity performance of I-JL-NWFET, assessed through parameters such as second and third-order transconductance (g_{m2}, g_{m3}), second and third-order input voltage intercept points (VIP2, VIP3), and third-order intermodulation distortion (IIP3 and IMD3) is improved at the higher temperature than that of JL-NWFET.

1. Introduction

The primary impetus behind the evolution of transistors towards miniaturization is to achieve higher density and enhanced performance for Integrated Circuit (IC) applications. However, this pursuit of miniaturization also introduces undesirable consequences in the form of adverse Short Channel Effects (SCEs) which impede technological progress [1]. To address these issues, the adoption of multi-gate transistors stands out as a promising solution for future technological advancements [2]. Nevertheless, the need for precise doping profiles at the metallurgical junctions and the necessity of a substantial thermal budget for dopant activation have become increasingly challenging as conventional Metal Oxide Semiconductor Field Effect Transistor (MOSFET) scaled down aggressively in size, which is also pertinent to future Complementary Metal Oxide Semiconductor (CMOS) technology [3].

Recently, Organic Thin-Film Transistors (OTFTs) have garnered significant attention for their commercial viability, primarily attributed to their mechanical flexibility, cost-effective fabrication, and semi-transparent characteristics [4]. Furthermore, leveraging the semiconductor properties of the organic materials used in their

© 2024 IOP Publishing Ltd

design, OTFTs find applications in diverse fields such as Organic Light Emitting Diode (OLED) [5, 6], digital inverters [7], organic solar cell [8], and more. Despite the widespread use of OTFT in various applications, there remains ample room for improvement in different aspects. The instability and degradation of organic materials stand out as primary concerns that need to be addressed for further advancements in these devices [9].

In response to these challenges, Junctionless Nanowire Field Effect Transistor (JL-NWFET) has been explored to replace the conventional MOSFET as well as OTFT. JL-NWFET provides an alternative approach by eliminating the necessity for p-n junctions at the source and drain ends, simplifying the manufacturing process and reducing the thermal budget [10]. However, to address issues such as the parasitic Bipolar Junction Transistor (BJT) phenomenon and insufficient volume depletion in JL-NWFET, the inclusion of an oppositely doped core in the channel region is explored [11, 12]. Furthermore, to overcome device degradation resulting from SCEs and parasitic BJT phenomena, as well as to enhance carrier transport efficiency, this paper introduces an Improved JL-NWFET (I-JL-NWFET) incorporating a Dual Material Gate (DMG) with a high-k material in the gate sidewall spacer region of the oppositely and equally doped core JL-NWFET.

The DMG structure employs two dissimilar gate work functions ensuring that the gate work function at the source side is higher than that at the drain side. This discrepancy in work function creates a potent electric field at the source, promoting the acceleration of electrons within the channel. Consequently, this enhanced acceleration increases drain current during the ON-state, thereby improving the current driving capability of the device. Moreover, the lower gate work function toward the drain reduces the magnitude of the drain-channel electric field experienced by electrons. As a result, the device is less susceptible to hot carrier-induced degradation and SCEs [13–15]. Additionally, the inclusion of the high-k material in the gate sidewall spacer region further alleviates the Lateral Band-to-Band Tunneling (L-BTBT) of electrons in the OFF-state [16, 17].

Recently, there has been a substantial increase in demand for portable electronic devices with high battery backup. To meet this heightened demand, it becomes imperative to increase transistor density and consequently scale the device's dimensions without compromising its performance. Moreover, elementary digital/analog circuits widely employed in sectors such as the military, automobile, nuclear, satellite communication, space, infrared detectors, and terrestrial systems are highly temperature-dependent [18]. The high-temperature reliability of a device ensures prolonged durability and stability. Additionally, linearity is a crucial requirement in all Radio Frequency (RF) systems, minimizing intermodulation and higher-order harmonics at the output. While there are system-level techniques for improving linearity, they often involve complex circuitry [19–21]. A transistor-level linearization is more suitable for power amplifiers in portable systems, necessitating an analysis of the linearity behavior at the device level. Therefore, it is crucial to investigate the impact of temperature variation on the device's RF, analog, and linearity performance for enhanced reliability. The investigation of temperature ranging from 200 K to 500 K is particularly relevant due to the diverse range of temperature-sensitive applications in electronics [22] and biomedicine [23]. Other reports are also available that have investigated the impact of temperature on similar ranges [24, 25].

In this paper, a detailed analysis of the scaling aspects of I-JL-NWFET and JL-NWFET is presented. Subsequently, an analysis of the electrical, analog/RF, and linearity performance is conducted, considering temperature variations ranging from 200 K to 500 K is also explored. The paper is structured into five sections, commencing with section 1 as the introduction. Section 2 outlines the device structure and characteristics, while section 3 delves into the impact of scaling on the electrical, analog/RF, and linearity performance of I-JL-NWFET and JL-NWFET. Additionally, section 4 scrutinizes the influence of temperature variations on the electrical, analog/RF, and linearity performance of I-JL-NWFET and JL-NWFET. The paper concludes with section 5, summarizing the findings and implications.

2. Device structures, specifications, and characteristics of JL-NWFET and I-JL-NWFET

For 3-D structure simulation of I-JL-NWFET and JL-NWFET, the SILVACO TCAD simulator is used [26]. Initially, to validate the simulation setup for conventional JL-NWFET, various physical models available in SILVACO, Technology Computer Aided Design (TCAD) has been employed and a combined transfer characteristics of simulation and experimental work [27] is plotted corresponding to $V_{DS} = 1$ V in figure 1. The Shockley-Read-Hall (SRH) recombination model is employed to address minority recombination, while the Bandgap Narrowing (BGN) model is utilized to factor in the impact of high doping in the channel. The inclusion of tunneling of electrons is achieved through the non-local BTBT model. The Lombardi mobility model (CVT), along with concentration and temperature-dependent and field mobility models, is also integrated. The convergence of results is facilitated by the numerical Newton method.

In figures 2(a) and (b), the 3-D and 2-D schematic diagrams of JL-NWFET are depicted, respectively. The structural parameters of JL-NWFET, including channel length (L), channel thickness (t_{si}), and gate oxide thickness (t_{ox}), adhere to the dimensions specified in the experimental paper [28], maintaining values of 60 nm,

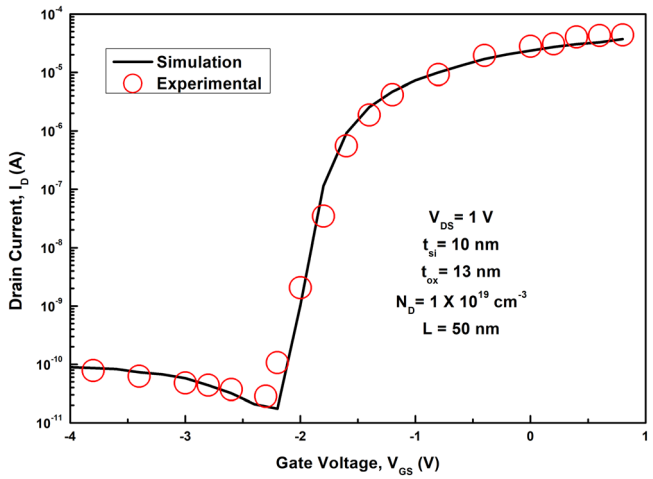


Figure 1. Combined plot of simulation and experimental results.

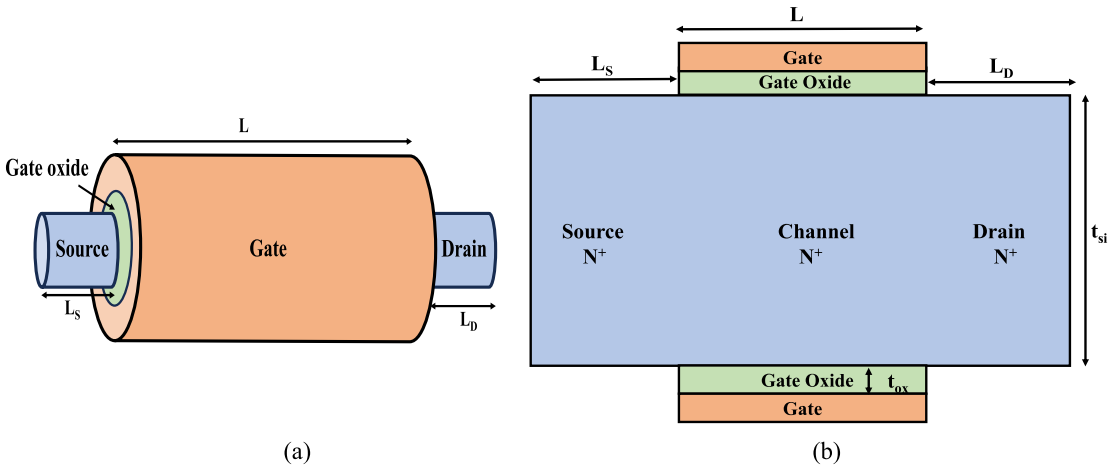


Figure 2. (a) 3-D and (b) 2-D cross-sectional view of JL-NWFET.

20 nm, and 1 nm, respectively. The length of source (L_S) and drain (L_D) is taken as 25 nm [11]. The SiO_2 is used as the gate oxide material. Subsequently, the consistent n-type doping concentration of 10^{19} cm^{-3} is utilized for defining the source, drain, and channel regions along with a high gate work function of 5.1 eV [11]. This heavy doping concentration also plays a pivotal role in ensuring the formation of Ohmic contacts at the metal-semiconductor interface [29]. By maintaining uniform doping levels at the source and drain regions, the possibility of Schottky contact formation can be mitigated.

Figures 3(a) and (b) showcase the 3-D and 2-D schematic depictions of I-JL-NWFET, respectively. I-JL-NWFET integrates a core that is equally and oppositely doped (p-doped), resulting in a reduced conduction layer referred to as the shell region (n-doped). The thickness of the core (t_{core}) and the shell (t_{shell}) contributes to t_{si} with $t_{si} = (2 \times t_{shell}) + t_{core}$. Furthermore, t_{core} is taken as 10 nm which is exactly half of t_{si}. The other structural parameters (L_G, t_{si}, t_{ox}, $L_{S/D}$) of the I-JL-NWFET mirror those of the JL-NWFET. The DMG structure in I-JL-NWFET comprises two gate metals (M_1 and M_2), each measuring 30 nm in length, with M_1 and M_2 having work functions of 5.1 eV (gold) and 4.6 eV (copper), respectively [30]. Additionally, HfO_2 is used as the gate sidewall spacer. Among the options explored, including air ($k = 1$), SiO_2 ($k = 3.9$), HfO_2 ($k = 22$) and Ta_2O_5 ($k = 26$), HfO_2 emerged as the most promising choice for enhancing device performance [12]. The device specifications of the JL-NWFET and I-JL-NWFET are outlined in table 1.

Figure 4 compares the transfer characteristics of JL-NWFET and I-JL-NWFET in linear and logarithm scales, respectively. The I-JL-NWFET demonstrates a remarkable enhancement in I_{ON}, surpassing that of the JL-NWFET by 23.1%. This improvement is credited to the emergence of an additional electric field at the interface of dissimilar gate metals within the DMG architecture of the I-JL-NWFET, thereby augmenting carrier velocity. Additionally, the incorporation of the high-k spacer proves instrumental in diminishing parasitic

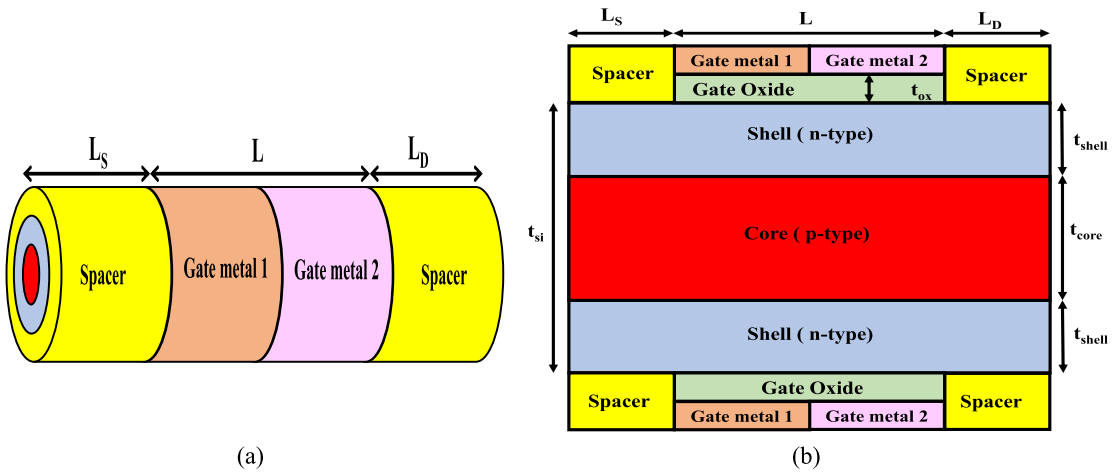


Figure 3. (a) 3-D and (b) 2-D cross-sectional view of I-JL-NWFET.

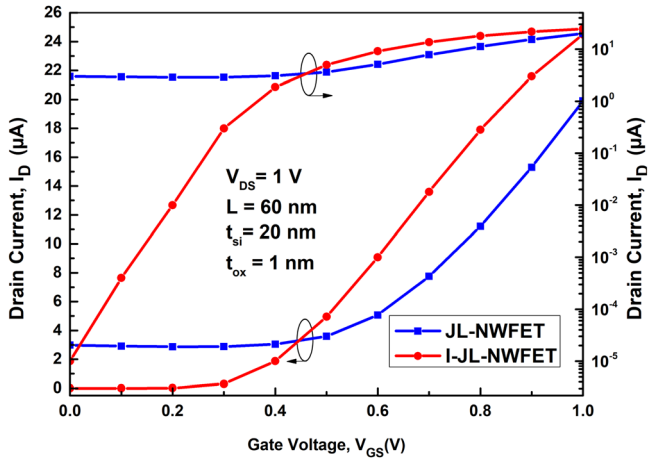


Figure 4. Plots of drain current of I-JL-NWFET and JL-NWFET in linear and log scale.

Table 1. Device specification for JL-NWFET and I-JL-NWFET.

Device parameter		JL-NWFET	I-JL-NWFET	
Channel	Material	Silicon	Silicon	
	Length, L (nm)	60	30 (L_{M1})	30 (L_{M2})
	Thickness, t_{si} (nm)	10	10	
	Doping, N_D (cm^{-3})	1×10^{19}	1×10^{19}	
Core	Material	—	Silicon	
	Thickness, t_{core}(nm)	—	10	
	Doping, N_A (cm^{-3})	—	1×10^{19}	
Gate	Material	Gold	Gold (M_1)	Copper (M_2)
	Workfunction (eV)	5.1	5.1 (M_1)	4.6 (M_2)
Gate Oxide	Material	SiO_2	SiO_2	
	Thickness, t_{ox}(nm)	1	1	
Source/Drain	Length, $L_{S/D}$ (nm)	25	25	
Spacer	Material	—	HfO_2	
	Length	—	25	

resistance beneath the spacer region, courtesy of the fringe field effect. The electrical performance of I-JL-NWFET is compared with JL-NWFET at the aforementioned structural parameters and results are succinctly presented in table 2.

Table 2. Performance comparison of JL-NWFET and I-JL-NWFET.

Parameter	V_T (V)	I_{ON} (μA)	I_{OFF} (μA)	I_{ON}/I_{OFF}	SS (mV/decade)
JL-NWFET	0.208	19.9	2.98	6.68	541.7
I-JL-NWFET	0.240	24.5	1.68×10^{-5}	1.81×10^{6}	64.8

Moreover, I_{OFF} is reduced by an order of 5 and accordingly, I_{ON}/I_{OFF} is improved by an order of 6 when compared to JL-NWFET. The integration of an oppositely doped core contributes to volume depletion, thereby mitigating I_{OFF}. The DMG architecture, in conjunction with the high-k spacer, widens the tunneling width and effectively suppresses the onset of the parasitic BJT phenomenon. Besides this, the Subthreshold Slope (SS) is improved by 88% in I-JL-NWFET as compared to JL-NWFET due to improved subthreshold characteristics exhibited by I-JL-NWFET. The most dominant parameters that contributes to the enhanced performance of I-JL-NWFET are the oppositely doped core and the DMG structure. As mentioned in the introduction section that conventional JL-NWFET suffers insufficient volume depletion and onset of parasitic BJT mechanism during OFF-state leading to high I_{OFF}. Moreover, JL-NWFET tends to experience reduced I_{ON} due to their heavily doped nature. The inclusion of an oppositely doped core facilitates hole recombination from the core and electrons in the channel, aiding in depletion during the OFF-state and consequently reducing I_{OFF}. However, it does not significantly affect I_{ON}. To address the need for increased I_{ON}, the DMG structure is introduced. This architecture amplifies the electric field at the source side by employing a higher gate work function at the source than at the drain, which accelerates electrons in the channel during the ON-state. This acceleration leads to enhanced I_D during the ON-state. The utilization of the high-k spacer, helps in reducing I_{OFF} by aiding in channel depletion beyond the gate region, is not as significant as the oppositely doped core and DMG structure in improving I-JL-NWFET performance.

3. Scaling impact on device performance

The device scaling represents a fundamental procedure within the semiconductor industry aimed at enhancing the transistor density on the chip. Therefore, it is imperative to assess how device scaling influences the performance of I-JL-NWFET. In this section, the impact of channel length (L) and channel thickness (t_{si}) scaling on the electrical, analog/RF and linearity parameters of I-JL-NWFET is analyzed. Furthermore, the results are also compared with that of conventional JL-NWFET.

3.1. Impact of channel length scaling on device performance

To observe the impact of channel length (L), an analysis is conducted on the electrical, analog/RF and linearity parameters of I-JL-NWFET and conventional JL-NWFET, with variations ranging from 70 nm to 20 nm keeping $t_{si} = 20$ nm, $L_{S/D} = 25$ nm, and $t_{ox} = 1$ nm as constant.

3.1.1. Impact of channel length scaling on electrical parameters

In this subsection, the impact of L on the electrical characteristics such as threshold voltage, drain current, OFF-state current, subthreshold slope and DIBL of I-JL-NWFET and JL-NWFET is analyzed.

Figure 5(a) illustrates the influence of L variation on the threshold voltage (V_T) and drain current (I_D) of I-JL-NWFET and JL-NWFET. The observed trend in figure 5(a) indicates a significant decrease in the V_T for conventional JL-NWFET when scaling down the L whereas I-JL-NWFET shows more consistent and lower variations in V_T. Notably, a sudden decline in V_T at an L of 20 nm suggests a loss of control of the gate voltage over the channel with a relatively large channel thickness. The drain current (I_D) for the scaled L is plotted in figure 5(a) at $V_{GS} = 1$ V. It corresponds to the maximum ON-state current of junctionless transistors as they are normally designed so that they operate in flat band condition in the ON-state [10].

The curve demonstrates that the I_D of both I-JL-NWFET and JL-NWFET increases with the scaling of L. The I_D of I-JL-NWFET surpasses that of JL-NWFET as the L is scaled down. Additionally, I-JL-NWFET shows enhancement of I_D due to the incorporation of DMG and high-k spacer implying better control over the flow of current through the channel.

The impact of L variation on the I_{OFF} and I_{ON}/I_{OFF} ratio of I-JL-NWFET and JL-NWFET is illustrated in figure 5(b). In the case of JL-NWFET, the I_{OFF} is higher than that of I-JL-NWFET as the L is scaled down. This is attributed to inadequate volume depletion in the large channel thickness and the initiation of parasitic BJT effects [31]. Typically, smaller L leads to higher I_{OFF}, which can be a concern for power efficiency and heat generation in integrated circuits. Conversely, I-JL-NWFET shows significantly better I_{OFF} variation at scaled L due to the presence of an opposite dope core which helps in volume depletion during OFF-state. Additionally,

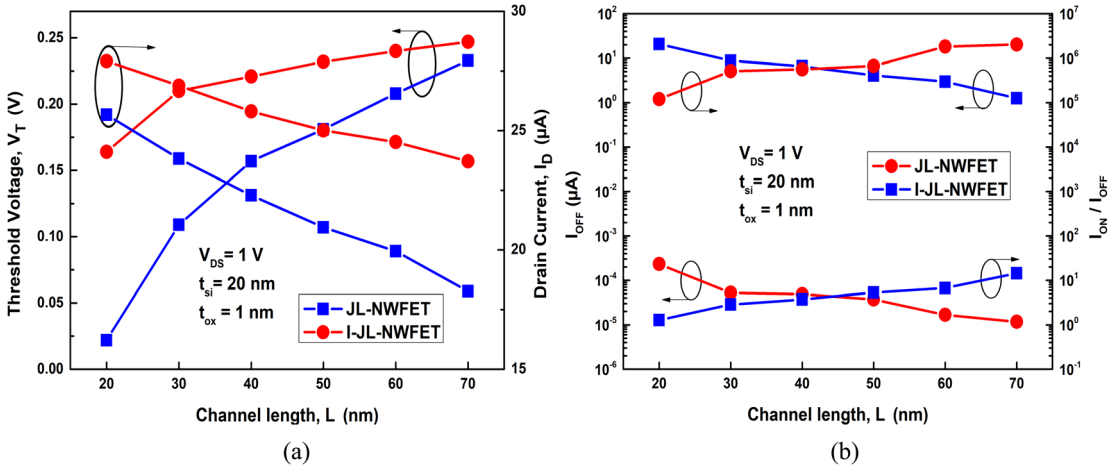


Figure 5. Plots of (a) threshold voltage and drain current (b) I_{OFF} and I_{ON}/I_{OFF} ratio of I-JL-NWFET and JL-NWFET for different channel lengths.

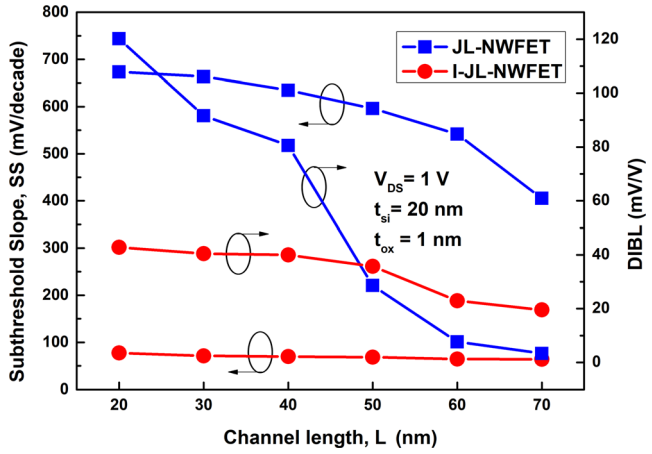


Figure 6. Plots of (a) SS and DIBL of I-JL-NWFET and JL-NWFET for different channel lengths.

DMG suppresses SCEs and Hot Carrier Effects (HCEs) more effectively and high-k material in the spacer region helps to deplete the channel beyond the gate edge. However, at a 20 nm channel length, there is a significant increase in the I_{OFF} of I-JL-NWFET, indicating poor electrostatic control of gate voltage. Consequently, the switching ratio (I_{ON}/I_{OFF}) of JL-NWFET and I-JL-NWFET decreases with decreasing L as portrayed in figure 5(b). Besides this, I-JL-NWFET depicts higher I_{ON}/I_{OFF} than JL-NWFET at scaled L. The I_{ON}/I_{OFF} ratio is determined by [10, 14]

$$\frac{I_{ON}}{I_{OFF}} = \frac{I_{DS}\,at\,V_{DS} = 1\ V\ and\ V_{GS} = 1\ V}{I_{DS}\ at\ V_{DS} = 1\ V\ and\ V_{GS} = 0\ V} \tag{1}$$

The Subthreshold Slope (SS) and Drain Induced Barrier Lowering (DIBL) are critical parameters when analyzing SCEs especially as L is scaled down. Figure 6 illustrates the influence of L variation on SS and DIBL. A lower SS is desirable to ensure efficient switching between the ON and OFF-states in the subthreshold region. SS is expressed as [10, 32]

$$SS = \frac{\partial V_{GS}}{\partial \log_{10}(I_D)} mV/decade \tag{2}$$

It can be noted from figure 6 that SS is lower for I-JL-NWFET than JL-NWFET. In JL-NWFET, the SS rises with the scaling of L because of an increase in I_{OFF}. This, in turn, leads to a reduction in the slope of the I_D-V_{GS} characteristics as the device transitions from the OFF-state to the ON-state. On the contrary, I-JL-NWFET shows less variation of SS at scaled L, implying improved subthreshold characteristics due to DMG and high-k spacer.

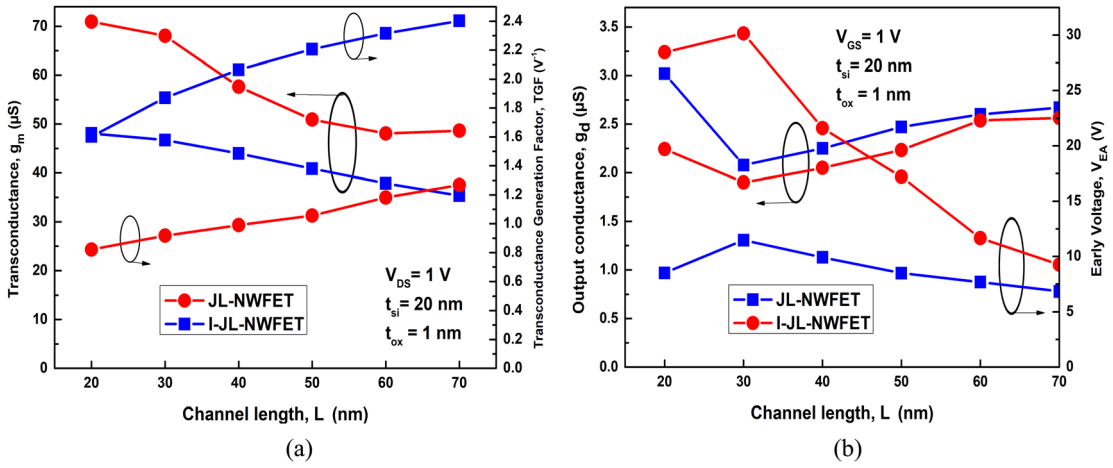


Figure 7. Plots of (a) transconductance and TGF (b) output conductance and early voltage of I-JL-NWFET and JL-NWFET for different channel lengths.

The DIBL refers to the reduction in the threshold voltage caused by an increase in the drain voltage and it is expressed as [33]

$$DIBL = \frac{V_T\ (V_{DS}\ = 0.05\ V) - V_T\ (V_{DS}\ = 1\ V)}{(V_{DS} = 1V) - (V_{DS} = 0.05V)} mV/V \tag{3}$$

The DIBL is quite severe in JL-NWFET as compared to I-JL-NWFET with scaling of L as demonstrated in figure 6. The reduction in DIBL effect in I-JL-NWFET is due to the presence of the oppositely doped core in the channel and DMG structure. The inclusion of an oppositely doped core decreased the active channel thickness of the I-JL-NWFET, improving gate control effectiveness and reducing the influence of drain voltage on the channel's electrostatics. As a result, the narrow conduction area led to a significant reduction in DIBL [34]. Additionally, the DMG structure aids in mitigating DIBL by providing a screening effect against drain voltage [33].

3.1.2. Impact of channel length scaling on analog/RF parameters

This subsection investigates the impact of L on the analog/RF parameters such as transconductance, Transconductance Generation Factor, output conductance, early voltage, total gate capacitance and cut-off frequency of the I-JL-NWFET and JL-NWFET.

The transconductance (g_m) stands out as a crucial analog metric in assessing devices, as it measures the relationship between the change in current and the variation in V_{GS}. A higher g_m signifies an inherent enhancement in amplification capacity. The Transconductance Generation Factor (TGF) holds significance in assessing the power needed to achieve high-speed concerning gate bias. TGF plays a pivotal role in determining how the current can be effectively utilized to attain a desired transconductance. The g_m and TGF are expressed as [17, 35]

$$g_m = \frac{\partial I_D}{\partial V_{GS}}\Big|_{V_{DS}=constant} \tag{4}$$

$$TGF = \frac{g_m}{I_D} \tag{5}$$

Figure 7(a) displays the influence of L variation on the g_m and TGF of I-JL-NWFET and JL-NWFET at $V_{GS} = 1$ V. With the scaling of L, the g_m of both I-JL-NWFET and JL-NWFET increases, with g_m of I-JL-NWFET exceeding that of JL-NWFET. This is attributed to the enhanced I_D in I-JL-NWFET than JL-NWFET as explained in figure 5(a). The TGF exhibits a greater decrease in I-JL-NWFET compared to JL-NWFET as L is scaled down.

The decrease in TGF is mainly attributed to the rising I_D as L decreases, which counteracts any potential improvement in g_m.

Generally, analog circuit devices are frequently operated in the saturation region. While I_D typically remains unaffected by V_{DS} in the saturation region, the impact of V_{DS} on channel electrostatics introduces variations in I_D, attributed to short-channel behavior. Hence, to understand the relationship between V_{DS} and I_D, output conductance (g_d) is evaluated. Additionally, the early voltage (V_{EA}) is crucial in the analysis and design of analog circuits, as it helps in determining the intrinsic gain of the device. The g_d and V_{EA} are formulated as follows [17]

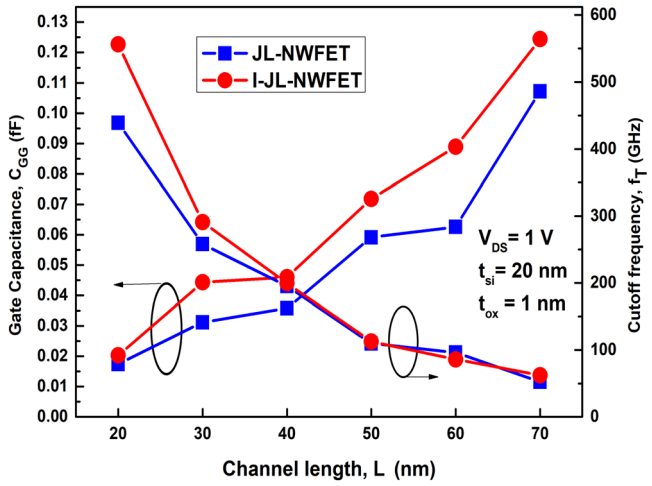


Figure 8. Plots of gate capacitance and cut-off frequency of I-JL-NWFET and JL-NWFET for different channel lengths.

$$g_d = \left.\frac{\partial I_D}{\partial V_{DS}}\right|_{V_{GS}=constant} \tag{6}$$

$$V_{EA} = \frac{I_D}{g_d} \tag{7}$$

The impact of L variation on the g_d and V_{EA} of I-JL-NWFET and JL-NWFET at $V_{DS} = 1$ V is portrayed in figure 7(b). The g_d of I-JL-NWFET reduces more than that of JL-NWFET as on scaling of L due to reduced DIBL in I-JL-NWFET than in JL-NWFET as observed in figure 6. Consequently, reduced g_d increases V_{EA} of I-JL-NWFET than that of JL-NWFET as on scaling of L. However, scaling of L below 30 nm may increase effect of drain bias leading to increase in g_d and decrease in V_{EA} in both I-JL-NWFET and JL-NWFET.

The total gate capacitance as parasitic capacitance (C_{GG}) and cut-off frequency (f_T) are important parameters for RF applications. C_{GG} in RF amplifiers may cause RF amplifiers to have low gain due to parasitic loss. The f_T is defined as the frequency at which the short circuit current gain drops to 0 dB. Mathematically, C_{GG} and f_T are formulated as [15]

$$C_{GG} = C_{GS} + C_{GD} \tag{8}$$

$$f_T = \frac{g_m}{2\pi C_{GG}} \tag{9}$$

where C_{GS} and C_{GD} are parasitic capacitances between gate-to-source and gate-to-drain terminals, respectively. Figure 8 shows the influence of varying L on the C_{GG} and f_T of I-JL-NWFET and JL-NWFET at $V_{GS} = 1$ V. The increase of C_{GG} of I-JL-NWFET exceeds that of JL-NWFET due to the presence of high-k spacer which increases capacitance at any given L. Moreover, the observed increase of f_T of I-JL-NWFET than that of JL-NWFET on scaling L is mainly due to increase in g_m as observed in figure 7(a).

3.1.3. Impact of channel length scaling on linearity parameters

The impact of scaling of L on linearity parameters of the I-JL-NWFET and JL-NWFET are analyzed in this subsection.

The fundamental parameters, g_{m2} and g_{m3}, are significantly influenced during high-frequency operation, leading to distortion and indicating nonlinearity. To enhance linearity performance, the peak values of g_{m2} and g_{m3} must be as low as possible. The g_{m2} and g_{m3} are expressed as [20]

$$g_{m2} = \frac{\partial^2 I_D}{\partial V_{GS}^2} \tag{10}$$

$$g_{m3} = \frac{\partial^3 I_D}{\partial V_{GS}^3} \tag{11}$$

Figure 9(a) depicts the influence of L variation on the peak value of g_{m2} and g_{m3} of I-JL-NWFET and JL-NWFET. The peak value of g_{m2} of I-JL-NWFET and JL-NWFET occurs at 0.6 V and 0.7 V, respectively while the peak value of g_{m3} of I-JL-NWFET and JL-NWFET occurs at 0.4 V and 0.5 V, respectively. The peak value of g_{m2} and g_{m3} increases on decreasing L due to less V_{GS} control on large channel thickness. Moreover, I-JL-NWFET

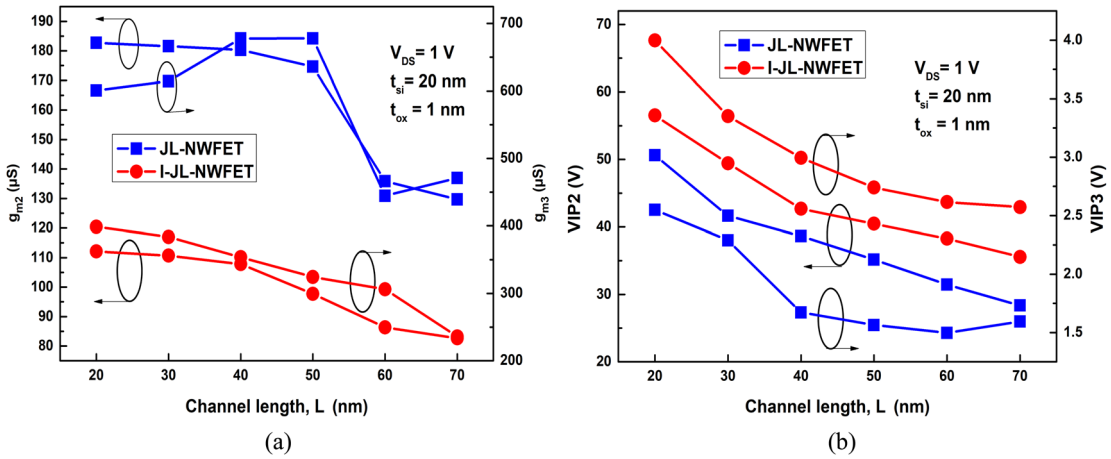


Figure 9. Plots of (a) g_{m2} and g_{m3} (b) VIP2 and VIP3 of I-JL-NWFET and JL-NWFET for different channel lengths.

exhibits low peak value of g_{m2} and g_{m3} than that of JL-NWFET at any given L, owing to inclusion of oppositely doped facilitating more effective V_{GS} control over channel region.

The set of metrics employed for evaluating distortion characteristics from DC parameters includes voltage intercept points (VIP2 and VIP3). The VIP2 signifies the extended input gate voltage where the voltage level of the fundamental signal equals that of the second-order harmonic signal. Beyond this crucial voltage, distortion is influenced by the amplitude of the first-order tone. The VIP3 represents the extrapolated input gate voltage at which the voltage levels of the fundamental and third-order harmonic signals become comparable. For high linearity performance and minimal distortion operation, a higher VIP2 and VIP3 is essential. These metrics can be expressed as follows [20]

$$VIP2 = 4 \times \frac{g_m}{g_{m2}} \tag{12}$$

$$VIP3 = \sqrt{24 \times \frac{g_m}{g_{m3}}} \tag{13}$$

Figure 9(b) illustrates the impact of L variation on the peak value of VIP2 and VIP3 of I-JL-NWFET and JL-NWFET. The peak value of VIP2 of I-JL-NWFET and JL-NWFET is observed at 0.7 V and 0.8 V, respectively while the peak value of VIP3 of I-JL-NWFET and JL-NWFET is observed at 0.6 V and 0.7 V, respectively. The peak value of VIP2 and VIP3 observed in I-JL-NWFET increases on decreasing L due enhancement in g_m that counteracts any increases in g_{m2} and g_{m3}.

The third-order input intercept point (IIP3) denotes the extrapolated input power level at which the powers of first and third-order harmonics become equal. IMD3, representing third-order intermodulation distortion, signifies the extrapolated intermodulation distortion power where the first and third-order intermodulation powers are equal. Equations (15) and (16) are employed to evaluate IIP3 and IMD3, respectively, with R_s assumed to be 50 Ω in analog and RF applications [20, 21]. Achieving a high IIP3 value and a low IMD3 value is desirable for enhancing linear characteristics.

$$IIP3 = \frac{2}{3} \times \frac{g_m}{g_{m3} \times R_S} \tag{14}$$

$$IMD3 = \left[\frac{9}{2} \times (VIP3)^2 \times g_{m3}\right]^2 \times R_S \tag{15}$$

Figure 10 demonstrates the impact of L variation on the peak value of IIP3 and IMD3 of I-JL-NWFET and JL-NWFET. The peak value of IIP3 of I-JL-NWFET and JL-NWFET is observed at 0.8 V and 0.9 V, respectively while the peak value of IMD3 of I-JL-NWFET and JL-NWFET is observed at 0.6 V and 0.7 V, respectively. Due to increasing g_m on decreasing L, the peak value of IIP3 in I-JL-NWFET increases than that in JL-NWFET. Furthermore, the peak value of IMD3 of I-JL-NWFET increases on decreasing L due to increasing g_{m3} and VIP3. I-JL-NWFET exhibits low peak value of IMD3 compared to JL-NWFET owing to low peak value of g_{m3} in I-JL-NWFET as observed in figure 9(a).

At 30 nm of L, I-JL-NWFET demonstrates optimal values for I_{OFF}, I_{ON}/I_{OFF}, and DIBL. Consequently, L of the I-JL-NWFET is fixed at 30 nm for further analysis in subsequent sections.

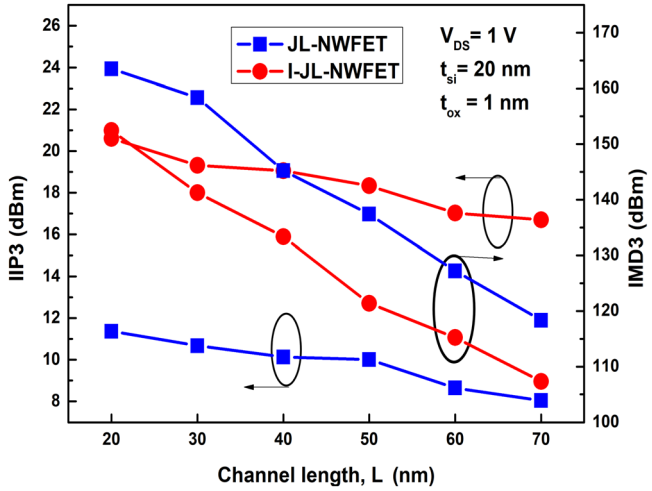


Figure 10. Plots of IIP3 and IMD3 of I-JL-NWFET and JL-NWFET for different channel lengths.

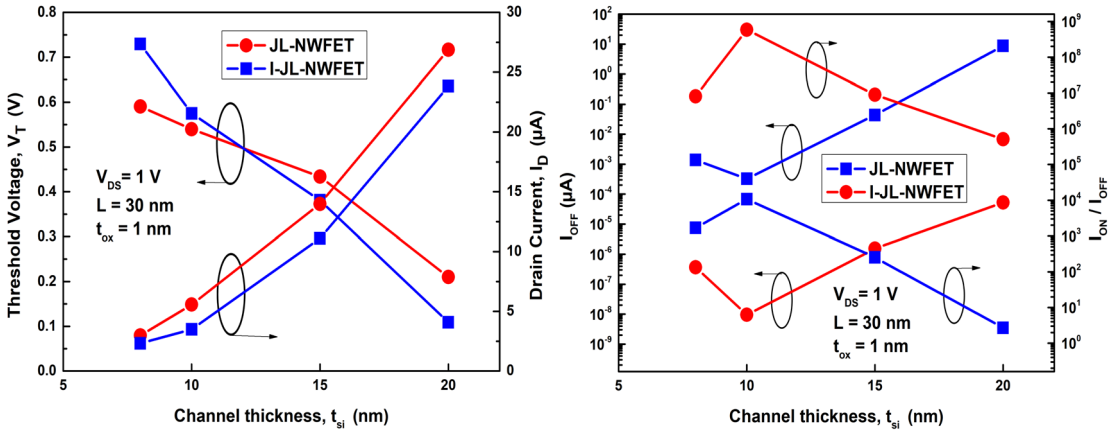


Figure 11. Plots of (a) threshold voltage and drain current (b) I_{OFF} and I_{ON}/I_{OFF} ratio of I-JL-NWFET and JL-NWFET for different channel thicknesses.

3.2. Impact of channel thickness scaling

The design of JL-NWFET is significantly influenced by the channel thickness (t_{si}), a crucial parameter essential for operation in various modes. Furthermore, the key concerns associated with JL-NWFET, namely insufficient volume depletion and leakage current, are strongly dependent on the t_{si} [27]. The impact of t_{si} on the electrical, analog/RF and linearity characteristics of both JL-NWFET and I-JL-NWFET is analyzed by varying it from 20 nm to 8 nm keeping $L = 30$ nm, $L_{S/D} = 25$ nm, and $t_{ox} = 1$ nm as constant. Moreover, t_{core} in I-JL-NWFET is varied as half of t_{si} during each scaling of t_{si}.

3.2.1. Impact of channel thickness scaling on electrical parameters

In this subsection, the analysis focuses on the impact of L on the electrical characteristics of I-JL-NWFET and JL-NWFET, including I_D, V_T, I_{OFF}, SS, and DIBL.

The impact of t_{si} scaling results in an increase in V_T for both JL-NWFET and I-JL-NWFET, as depicted in figure 11(a). In the case of large t_{si} of JL-NWFET, there is insufficient volume depletion, requiring less gate voltage to turn the transistor ON. I-JL-NWFET exhibits less variation in V_T as t_{si} is scaled, compared to JL-NWFET because of the presence of an opposite doped core that reduces active shell thickness for conduction.

Furthermore, as shown in figure 11(a), I_D decreases with a reduction in t_{si} for both JL-NWFET and I-JL-NWFET. In the case of JL-NWFET, where bulk conduction occurs, the decrease in cross-sectional area for current flow leads to a decrease in I_D. Conversely, in I-JL-NWFET, which features a DMG, the presence of dual materials enhances carrier mobility. This results in an increase in I_D compared to JL-NWFET, despite the

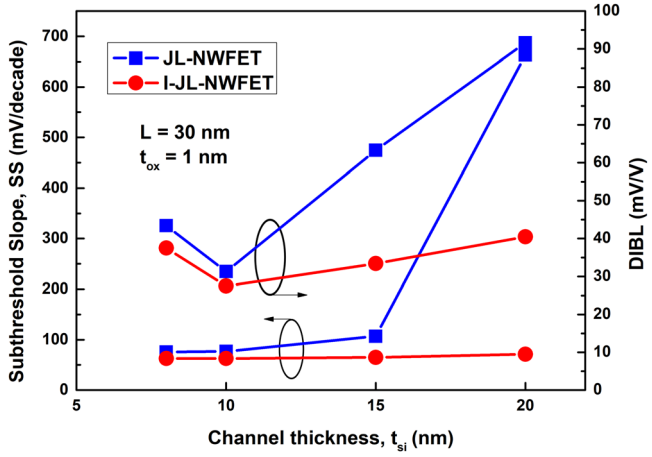


Figure 12. Plots of SS and DIBL of I-JL-NWFET and JL-NWFET for different channel thicknesses.

Table 3. Performance comparison of I-JL-NWFET with JL-NWFET and other reported devices.

Parameters	[10]	Re-S/D-JL-GAA MOSFET [36]	DPGAA MOSFET [37]	JL-NWFET	I-JL-NWFET
I_{OFF} (A)	1×10^{-15}	9.5×10^{-10}	5.34×10^{-13}	3.25×10^{-10}	9.55×10^{-15}
I_{ON}/I_{OFF}	1×10^{6}	1.2×10^{4}	2.14×10^{8}	1.07×10^{4}	5.83×10^{8}
SS (mV/decade)	64	73.63	69.96	76.83	62.75
DIBL (mV/V)	—	49.47	—	31.30	27.51

presence of the opposite doped core that reduces the active shell thickness for conduction. Figure 11(b) illustrate the influence of t_{si} variation on the I_{OFF} and I_{ON}/I_{OFF} ratio. The higher I_{OFF} in JL-NWFET with a t_{si} of 20 nm is attributed to insufficient volume depletion and the initiation of the parasitic BJT phenomenon. This effect diminishes with the scaling of t_{si}, allowing the gate voltage to more effectively control the channel. However, in I-JL-NWFET, the presence of an oppositely doped core aids in carrier depletion in the OFF-state. Additionally, the utilization of a DMG, including high-k material in the spacer region, further contributes to reducing I_{OFF} compared to JL-NWFET on the scaling of t_{si}. Both JL-NWFET and I-JL-NWFET experience an increase in I_{OFF} at a t_{si} of 8 nm due to a rise in SCEs. However, it is worth emphasizing that the I_{OFF} recorded for I-JL-NWFET at a t_{si} of 8 nm is comparatively lower than that observed in JL-NWFET under the same conditions. Consequently, I_{ON}/I_{OFF} ratio in I-JL-NWFET is higher than JL-NWFET.

The impact of varying t_{si} on SS and DIBL is shown in figure 12. The SS in JL-NWFET decreases drastically when t_{si} is scaled due to the significant reduction in the OFF-state current. Additionally, SS also increases since scaling of t_{si} beyond 10 nm increases I_{OFF} in JL-NWFET. However, SS in I-JL-NWFET shows less sensitivity towards scaled t_{si} due to improved I_{OFF} as discussed in figure 11(b). Furthermore, decreasing the nanowire t_{si} improves the efficacy of gate control, leading to a diminished impact of the drain voltage on the electrostatics in the channel region. Consequently, the effective gate control provided by the smaller t_{si} markedly reduces DIBL in JL-NWFET till 10 nm of t_{si} beyond which it increases due to reduced control of gate voltage over the channel region, as illustrated in figure 12. Nevertheless, in the case of I-JL-NWFET, DIBL exhibits less sensitivity on scaling t_{si}. This is attributed to the presence of the core with opposite doping, which effectively reduces the active channel region, and DMG, which improves gate control on the reduced active channel region.

Table 3 compares the electrical performance of I-JL-NWFET with JL-NWFET, experimental work [10], Recessed Source/Drain Junctionless Gate All Around (Re-S/D-JL-GAA) MOSFET [36] and Dielectric Pocket Gate All Around (DPGAA) MOSFET [37]. It can be observed that I-JL-NWFET showcases higher I_{ON}/I_{OFF}, and less value of I_{OFF}, SS and DIBL when compared to other published work. The superior electrical performance demonstrated by the I-JL-NWFET underscores its viability for various electronic applications.

3.2.2. Impact of channel thickness scaling on analog/RF parameters

The impact of scaling of t_{si} on analog/RF parameters such as g_m, TGF, g_d, V_{EA}, C_{GG} and f_T of the I-JL-NWFET and JL-NWFET are analyzed in this subsection.

Figure 13(a) illustrates the variation of t_{si} on g_m and TGF of I-JL-NWFET and JL-NWFET at $V_{GS} = 1$ V. The g_m of I-JL-NWFET and JL-NWFET decreases on scaling of t_{si} due to reduction of $I_{D.}$ Moreover, g_m of

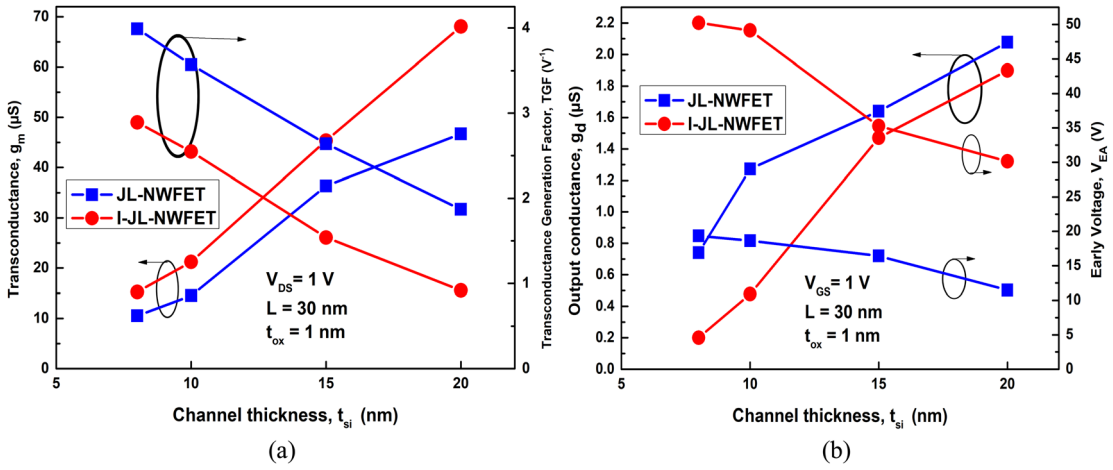


Figure 13. Plots of (a) transconductance and TGF (b) output conductance and early voltage of I-JL-NWFET and JL-NWFET for different channel thicknesses.

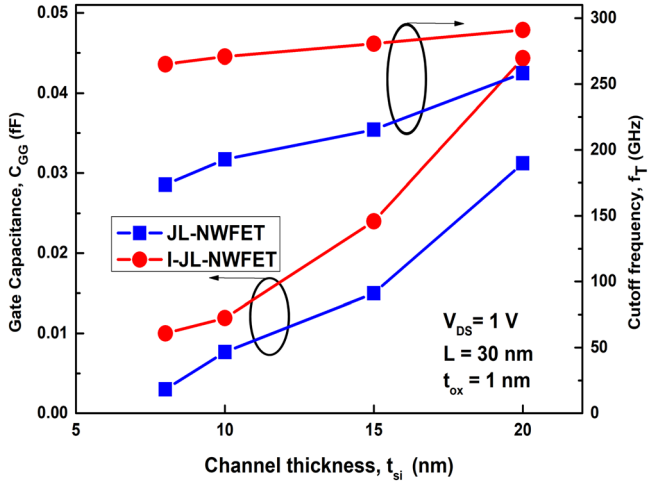


Figure 14. Plots of gate capacitance and cut-off frequency of I-JL-NWFET and JL-NWFET for different channel thicknesses.

I-JL-NWFET surpasses than that of JL-NWFET attributed to enhanced I_D at any given t_{si}, as exhibited in figure 11(a). Consequently, TGF of I-JL-NWFET and JL-NWFET increases as t_{si} is scaled down due to reduced I_D. Additionally, I-JL-NWFET exhibits higher TGF compared to JL-NWFET at any given t_{si}, due to DMG structure that increases carrier mobility in the channel.

Figure 13(b) illustrates g_d and V_{EA} of I-JL-NWFET and JL-NWFET at $V_{DS} = 1$ V on varying t_{si}. The g_d of I-JL-NWFET and JL-NWFET decreases on scaling of t_{si} due to reduction of I_D as observed in figure 11(a). Moreover, g_d of I-JL-NWFET is less compared to JL-NWFET at any given t_{si}, attributed due to the enhanced control of V_{GS} compared to V_{DS} on smaller channel thickness. Consequently, V_{EA} of I-JL-NWFET surpasses that of JL-NWFET due to lower g_d at any given t_{si}. It is also observed that there is marginal reduction of g_d below 10 nm of t_{si} leading to minimal change in V_{EA} of I-JL-NWFET.

Figure 14 depicts the impact of variation of t_{si} on C_{GG} and f_T of I-JL-NWFET and JL-NWFET at $V_{GS} = 1$ V. At any given t_{si}, I-JL-NWFET exhibit higher C_{GG} compared to JL-NWFET, primarily because of the inclusion of the high-k spacer, which enhances capacitance between gate to source and gate to drain. Furthermore, the noticeable enhancement in f_T of I-JL-NWFET compared to JL-NWFET on scaling t_{si} is predominantly attributed to the increase in g_m, as depicted in figure 13(a).

3.2.3. Impact of channel thickness scaling on linearity parameters

The analysis of this subsection focuses on the impact of variation of t_{si} on the linearity parameters of I-JL-NWFET and JL-NWFET, consisting g_{m2}, g_{m3}, VIP2, VIP3, IIP3 and IMD3.

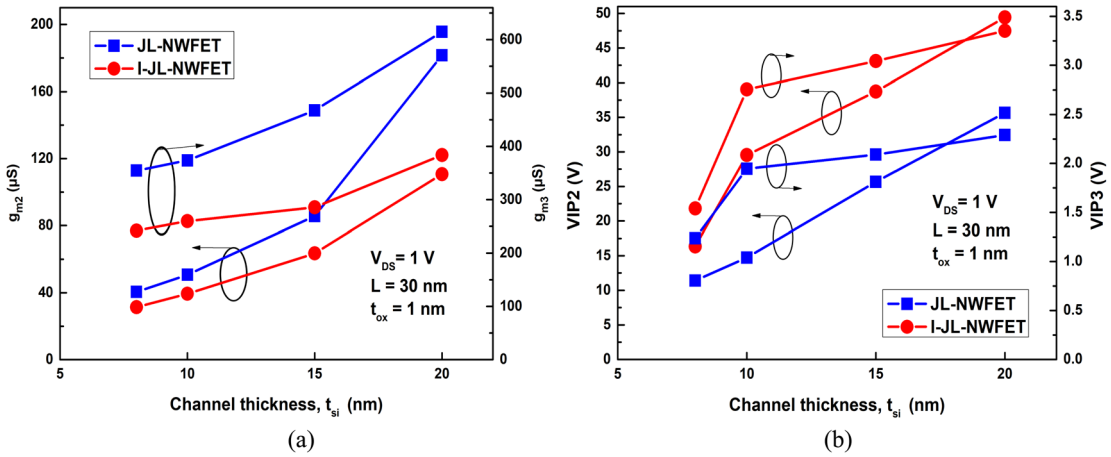


Figure 15. Plots of (a) g_{m2} and g_{m3} (b) VIP2 and VIP3 of I-JL-NWFET and JL-NWFET for different channel thicknesses.

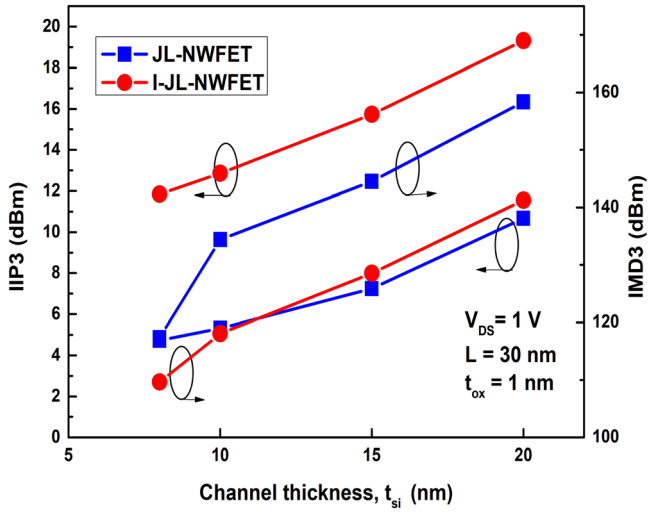


Figure 16. Plots of IIP3 and IMD3 of I-JL-NWFET and JL-NWFET for different channel thicknesses.

Figure 15(a) demonstrates the influence of t_{si} variation on the peak value of g_{m2} and g_{m3} of I-JL-NWFET and JL-NWFET. The peak value of g_{m2} of I-JL-NWFET and JL-NWFET occurs at 0.6 V and 0.7 V, respectively while the peak value of g_{m3} of I-JL-NWFET and JL-NWFET is observed at 0.4 V and 0.5 V, respectively. On decreasing t_{si}, I-JL-NWFET exhibits lower peak value of g_{m2} and g_{m3} compared to JL-NWFET due to enhanced V_{GS} control over smaller channel leading to reduced harmonic distortion [20].

Figure 15(b) shows the impact of varying t_{si} on the peak value of VIP2 and VIP3 of I-JL-NWFET and JL-NWFET. The peak value of VIP2 for I-JL-NWFET occurs at 0.7 V, whereas for JL-NWFET, it occurs at 0.8 V. Similarly, the peak value of VIP3 for I-JL-NWFET is observed at 0.6 V, while for JL-NWFET, it is observed at 0.7 V. I-JL-NWFET displays higher peak value of VIP2 and VIP3 compared to JL-NWFET due to lower peak value of g_{m2} and g_{m3}.

Figure 16 illustrates the impact of t_{si} variation on the peak value of IIP3 and IMD3 of I-JL-NWFET and JL-NWFET. The peak value of IIP3 of I-JL-NWFET and JL-NWFET is observed at 0.8 V and 0.9 V, respectively while the peak value of IMD3 of I-JL-NWFET and JL-NWFET is observed at 0.6 V and 0.7 V, respectively. Due to decreasing g_m on reducing t_{si}, the peak value of IIP3 in JL-NWFET reduces more than that in I-JL-NWFET. Furthermore, the peak value of IMD3 of I-JL-NWFET decreases on reducing t_{si} due to reducing VIP3. Notably, I-JL-NWFET demonstrates a lower peak value of IMD3 compared to JL-NWFET at any given t_{si}, owing to low peak value of g_{m3} in I-JL-NWFET as observed in figure 15(a).

I-JL-NWFET exhibits optimum value of I_{OFF}, I_{ON}/I_{OFF} and DIBL at 30 nm of L and 10 nm of t_{si}. Moreover, analog/RF and linearity parameters depend on electrical parameters. Therefore, the L and t_{si} of I-JL-NWFET are optimized at 30 nm and 10 nm, respectively for subsequent sections.

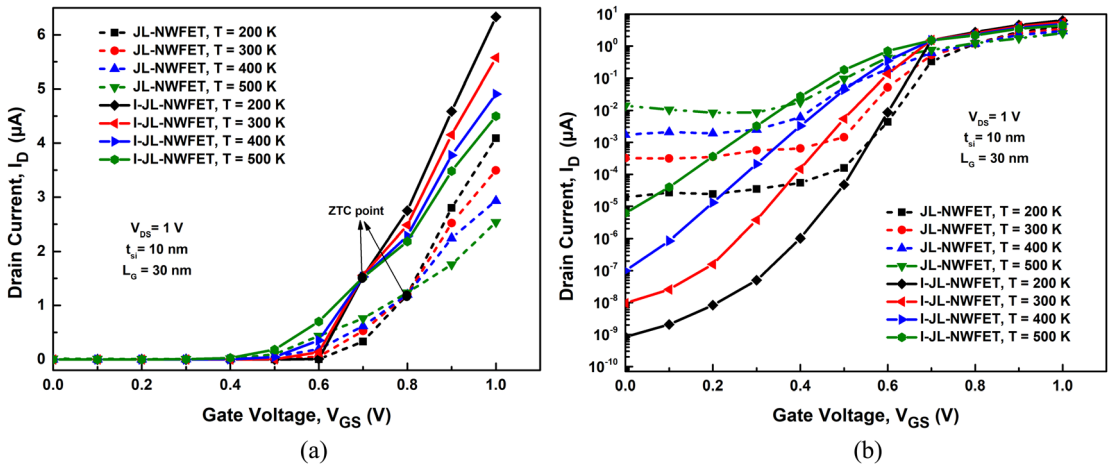


Figure 17. Transfer characteristics of I-JL-NWFET and JL-NWFET for different temperatures in (a) linear scale and (b) log scale.

4. Temperature impact on device performance

The temperature dependence analysis of nano-scale devices offers a more profound understanding of the intrinsic performance of the device, allowing for insights into its performance under varying temperature conditions. This section delves into exploring the impact of temperature on the characteristics of electrical, analog/RF, and linearity parameters in I-JL-NWFET and JL-NWFET.

4.1. Impact of temperature on electrical parameters

In this subsection, the impact of temperature on I-JL-NWFET and JL-NWFET is analyzed on the electrical characteristics such as drain current, threshold voltage, OFF-state current, subthreshold slope and DIBL by varying the temperature from 200 K to 500 K.

The dependence of temperature on transfer characteristics of I-JL-NWFET and JL-NWFET in linear and log scales are illustrated in figures 17(a) and (b), respectively. As the temperature rises, I_D initially rises at a specific V_{GS} and subsequently declines with further temperature increases. This point is called the Zero Temperature Coefficient (ZTC) at which I_D demonstrates almost no dependence on temperature [38]. Below the ZTC point, an increase in temperature leads to a decrease in V_T, causing an increase in I_D. Conversely, beyond the ZTC point, I_D decreases due to a decline in electron mobility with rising temperature. The ZTC of I-JL-NWFET and JL-NWFET are observed at 0.7 V and 0.8 V, respectively. In JL-NWFET, the primary degradation mechanism for electron mobility is phonon scattering, and this effect varies proportionally with $T^{-3/2}$ [38]. However, the I_D ($V_{GS} = 1$ V) of I-JL-NWFET is comparatively high in comparison to JL-NWFET attributed to the incorporation of DMG that enhances the current driving capability in the device as observed in figure 17(a). Moreover, it is found that the variation of I_D at $V_{GS} = 1$ V in I-JL-NWFET and JL-NWFET is 29% and 31.8%, respectively with an increase in temperature from 200 K to 500 K. This underscores the comparatively lower sensitivity of I-JL-NWFET to temperature variations when compared to JL-NWFET.

Additionally, I_{OFF} ($V_{GS} = 0$ V) increases in both I-JL-NWFET and JL-NWFET with an increase in temperature from 200 K to 500 K as depicted in figure 17(b). This is attributed to the increases in intrinsic carrier concentration (n_i) which is a dependent factor of temperature and expressed as [39]

$$n_i = \exp\left[\frac{-E_g}{2KT}\right] \tag{16}$$

where 'K' is the Boltzmann constant, E_g is bandgap of semiconductor, and 'T' is the absolute temperature. The I_{OFF} in I-JL-NWFET is comparable less in comparison to JL-NWFET due to an oppositely doped core which aids in depleting the channel during the OFF state.

The variation of Conduction Band (CB) energy and Valance Band (VB) energy along the device length of I-JL-NWFET and JL-NWFET for different temperatures at $V_{GS} = 0$ V is plotted in figure 18(a). As the band gap profile is analyzed below ZTC point of I-JL-NWFET and JL-NWFET, with an increase in temperature, band gap decreases that consequently increases I_D. Moreover, at any given temperature, JL-NWFET exhibits the smallest tunneling width from VB of channel to CB of drain region in comparison to I-JL-NWFET during OFF-state. This mechanism causes parasitic BJT and increases I_{OFF} in JL-NWFET [11]. However, the inclusion of an oppositely doped core widens this tunneling width and enhances source to channel barrier height that

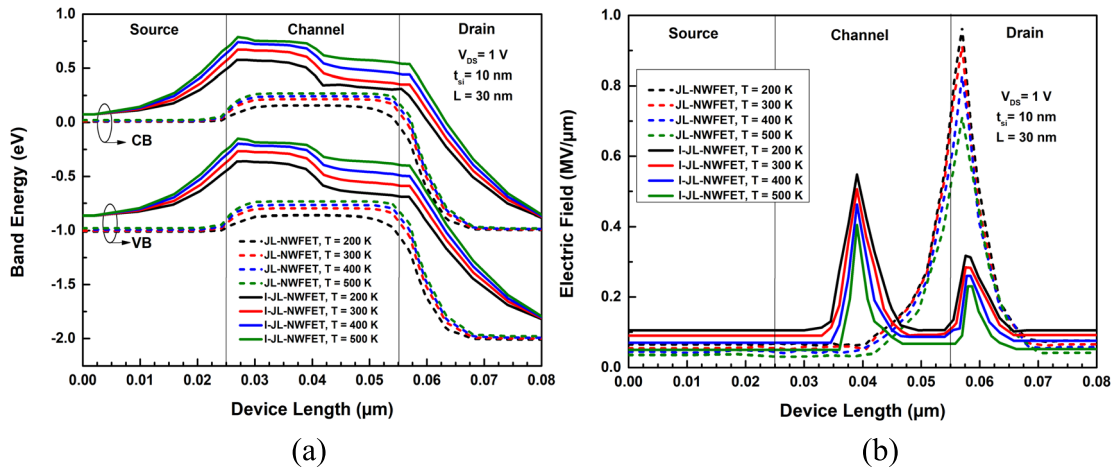


Figure 18. Plots of (a) energy band and (b) electric field of I-JL-NWFET and JL-NWFET for different temperatures.

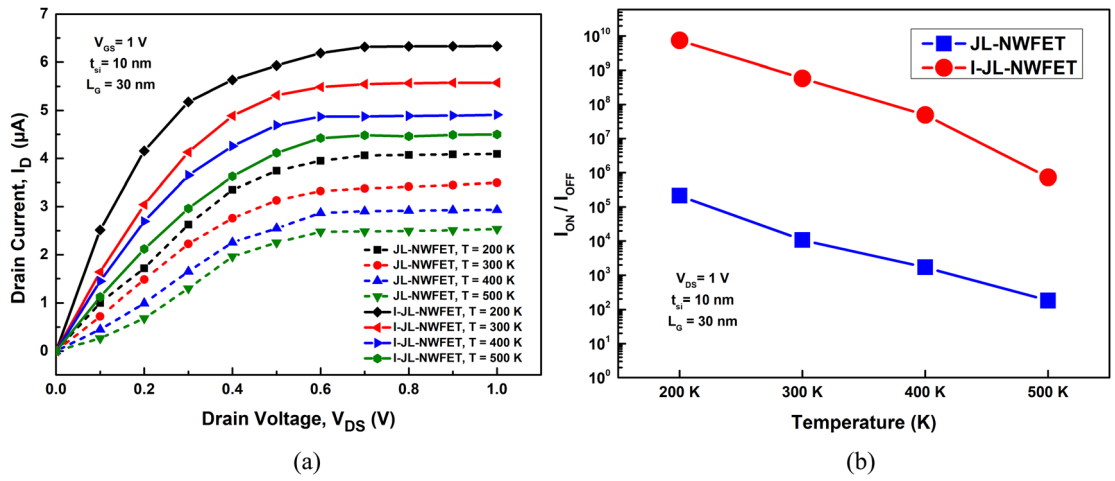


Figure 19. Plots of (a) output characteristics and (b) I_{ON}/I_{OFF} ratio of I-JL-NWFET and JL-NWFET for different temperatures.

cumulatively decreases I_{OFF} in I-JL-NWFET [11]. Additionally, the employment of two dissimilar gate work functions in the DMG structure results in an abrupt transition, further amplifying the tunneling width of I-JL-NWFET [13].

Figure 18(b) depicts the variation of the electric field along the device length of I-JL-NWFET and JL-NWFET for different temperatures at $V_{GS} = 1$ V. As the electric field is analyzed above ZTC point of I-JL-NWFET and JL-NWFET, with an increase in temperature, electric decreases, subsequently reducing I_D. At any particular temperature, I-JL-NWFET displays two electric field peaks leading to accelerating electrons in the channel, which consequently enhances I_D as compared to JL-NWFET in ON-state. Furthermore, the peak value of an electric field near drain region in I-JL-NWFET is reduced than that in JL-NWFET due to the lower work function of M_2. This characteristic enables better modulation of the V_{GS} over the channel rather than the V_{DS}, consequently mitigating SCE such as DIBL [14].

Figure 19(a) presents output characteristics variation of I-JL-NWFET and JL-NWFET with temperature at $V_{GS} = 1$ V. The I_D of I-JL-NWFET is observed to be higher than that of JL-NWFET. Analysis from figure 19(a) reveals that the ZTC for I-JL-NWFET and JL-NWFET occurs at 0.7 V and 0.8 V, respectively. Beyond this ZTC point, I_D decreases with increasing temperature. Consequently, as the output characteristics are plotted beyond their respective ZTC points, the I_D of both devices decreases as temperature rises. Moreover, for any given temperature, the I_D of I-JL-NWFET is observed to be higher than that of JL-NWFET as explained in figure 18(b).

The I_{ON}/I_{OFF} ratio of I-JL-NWFET and JL-NWFET at various temperatures is plotted in figure 19(b). As discussed in figure 17(b), an increase in temperature increases I_{OFF}, consequently decreasing I_{ON}/I_{OFF} ratio in both I-JL-NWFET and JL-NWFET. However, the I_{ON}/I_{OFF} ratio of I-JL-NWFET is notably higher compared to JL-NWFET. The reason behind this is the utilization of DMG architecture and oppositely doped core in

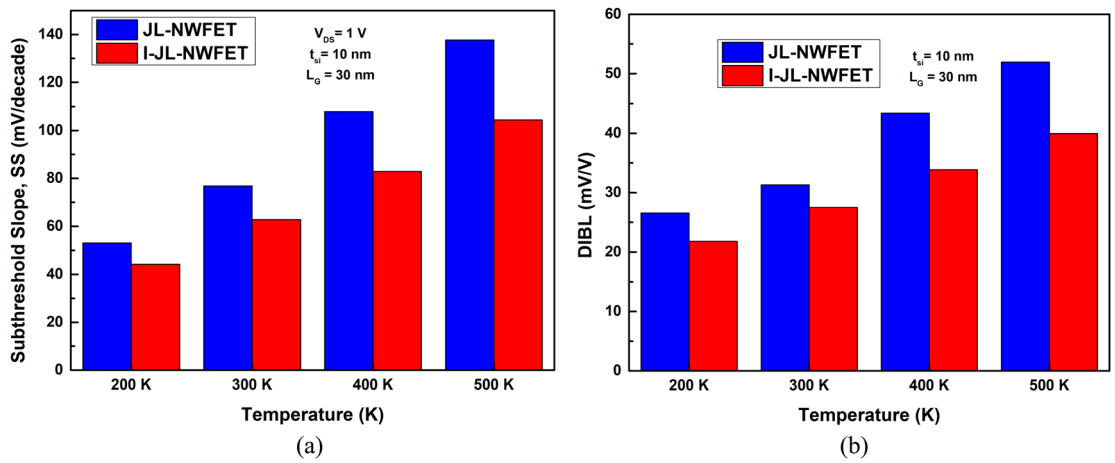


Figure 20. Plots of (a) SS and (b) DIBL of I-JL-NWFET and JL-NWFET for different temperatures.

Table 4. Effect of temperature on the electrical performance of JL-NWFET and I-JL-NWFET.

Parameters	Devices	200 K	300 K	400 K	500 K
I_{ON}/I_{OFF}	JL-NWFET	2.13×10^5	1.07×10^4	1.70×10^3	1.81×10^2
	I-JL-NWFET	7.40×10^9	5.83×10^8	4.93×10^7	7.37×10^5
SS (mV/decade)	JL-NWFET	53.14	76.83	107.85	137.68
	I-JL-NWFET	44.19	62.75	82.84	104.41
DIBL (mV/V)	JL-NWFET	26.54	31.30	43.40	51.95
	I-JL-NWFET	21.84	27.51	33.84	39.95

I-JL-NWFET, which work synergistically to increase I_{ON} and decrease I_{OFF}, respectively. The variation of SS and DIBL in I-JL-NWFET and JL-NWFET against increasing temperature is illustrated in figures 20(a) and (b), respectively. Notably, the observed variations in SS and DIBL are 29% and 28% less, respectively, in I-JL-NWFET compared to JL-NWFET when the temperature is increased from 200 K to 500 K. This indicates that I-JL-NWFET demonstrates a relatively lower sensitivity to temperature-induced changes in SS and DIBL compared to JL-NWFET over the specified temperature range.

The SS and DIBL in I-JL-NWFEET vary in a controlled manner with increasing temperature in comparison to JL-NWFET. At any given temperature, the reduced SS observed in I-JL-NWFET is credited to its high I_{ON}/I_{OFF} ratio, indicating superior switching speed compared to JL-NWFET. Furthermore, the DIBL of I-JL-NWFET is less than that of JL-NWFET at a specified temperature range attributed to the presence of the DMG architecture. As discussed in figure 18(b), the lower gate work function at the drain side than the source side results in a reduced electric field, allowing for greater control of the V_{GS} over the channel rather than the V_{DS}.

Table 4 provides a comprehensive comparison of the electrical performance of the I-JL-NWFET with that of JL-NWFET at various temperatures. Notably, the comparison demonstrates that parameters such as I_{ON}/I_{OFF} ratio, SS and DIBL are less susceptible to temperature fluctuations in the I-JL-NWFET when compared to the JL-NWFET.

4.2. Impact of temperature on analog/RF parameters

The analog/RF circuits are often sensitive to changes in temperature. An in-depth analysis of temperature variation is vital for ensuring the reliable and consistent performance of the device. This sub-section investigates the impact of temperature on the analog/RF parameters such as g_m, TGF, g_d, V_{EA}, C_{GG}, and f_T of the I-JL-NWFET and JL-NWFET.

Figures 21(a) and (b) exhibit the g_m and TGF variation with respect to V_{GS} at various temperatures, respectively. It is noted from figure 21(a) that the g_m of both I-JL-NWFET and JL-NWFET experiences an increase with temperature below ZTC. However, above the ZTC, the trends reverse, possibly attributed to an increase in the rate of scattering events with rising temperatures, as discussed in figure 17(a). Furthermore, for a given temperature, the g_m of I-JL-NWFET surpasses that of JL-NWFET owing to DMG architecture. Additionally, there is a decrease of 24.4% in g_m for I-JL-NWFET and 26.7% for JL-NWFET at $V_{GS} = 1$ V.

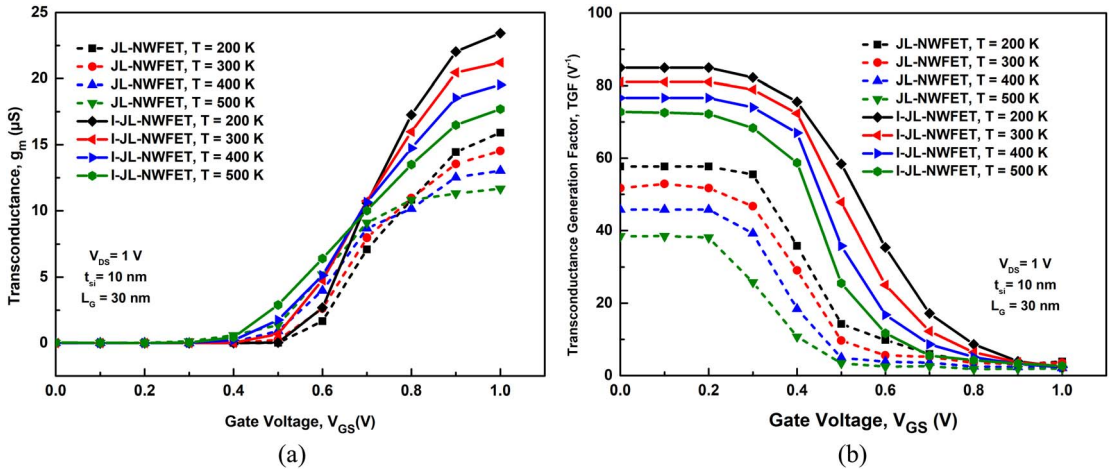


Figure 21. Plots of (a) transconductance and (b) TGF of I-JL-NWFET and JL-NWFET for different temperatures.

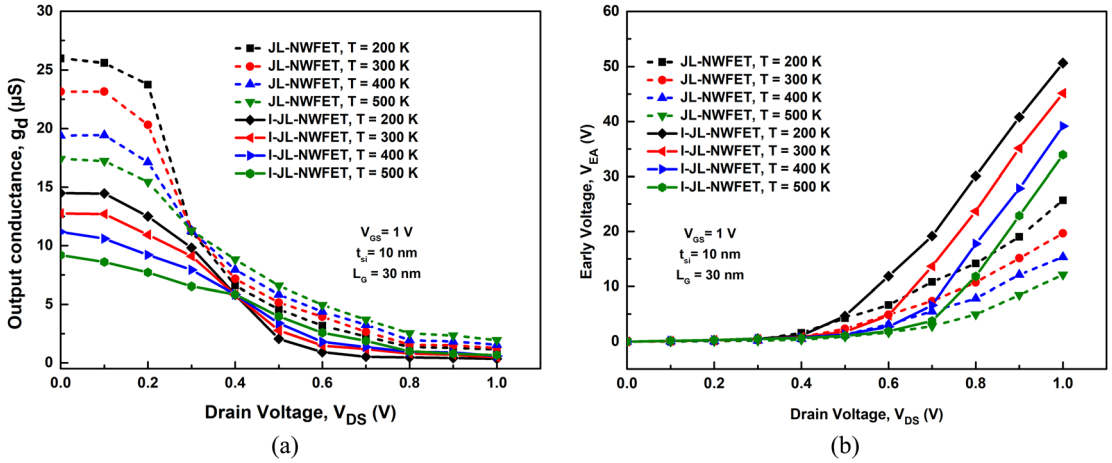


Figure 22. Plots of (a) output conductance and (b) early voltage of I-JL-NWFET and JL-NWFET for different temperatures.

Figure 21(b) reveals that in the subthreshold region, the TGF decreases by 14.32% in I-JL-NWFET and 33.40% in JL-NWFET, with an increase in temperature from 200 K to 500 K at $V_{GS} = 0$ V.

The impact of temperature on g_d and V_{EA} of I-JL-NWFET and JL-NWFET are portrayed in figures 22(a) and (b), respectively. At lower V_{DS}, the g_d is primarily influenced by Channel Length Modulation (CLM). Conversely, at higher V_{DS}, the g_d is governed DIBL [17]. CLM represents a SCE whereby the channel length is effectively shortened due to the application of drain bias. This phenomenon is particularly relevant because the output resistance of the device is directly proportional to the channel length. Consequently, when the channel length is shortened due to drain bias, it increases the output conductance.

As depicted in figure 22(a), it is evident that, at any given temperature, g_d in I-JL-NWFET is lower than JL-NWFET at higher V_{DS}. Additionally, a temperature increase at lower V_{DS} results in a reduction of g_d until a specific V_{DS}, beyond which g_d reverses its trend due to the occurrence of ZTC point and rise in DIBL with the temperature increment. Due to the lowest g_d of I-JL-NWFET at higher V_{DS} at any particular temperature, it attains a superior V_{EA} than that in JL-NWFET as illustrated in figure 22(b).

The enhanced V_{EA} contributes to an improved intrinsic gain of the device, particularly beneficial for analog amplifiers. Moreover, with the rise in temperature from 200 K to 500 K, V_{EA} decreases 32.9% in I-JL-NWFET and 52.72% in JL-NWFET.

Figures 23(a) and (b) depict the C_{GG} and f_T as a function of V_{GS} for temperatures ranging from 200 K to 500 K for I-JL-NWFET and JL-NWFET, respectively. From figure 23(a), it is perceived that, under low V_{GS}, C_{GG} is predominantly influenced by the depletion capacitance (C_{dep}) [38]. With the rising temperature, the surface potential diminishes, causing an enhancement in C_{dep}. Conversely, under high V_{GS} conditions, an increase in temperature leads to a reduction in C_{GG}. Notably, I-JL-NWFET exhibits a higher C_{GG} compared to JL-NWFET

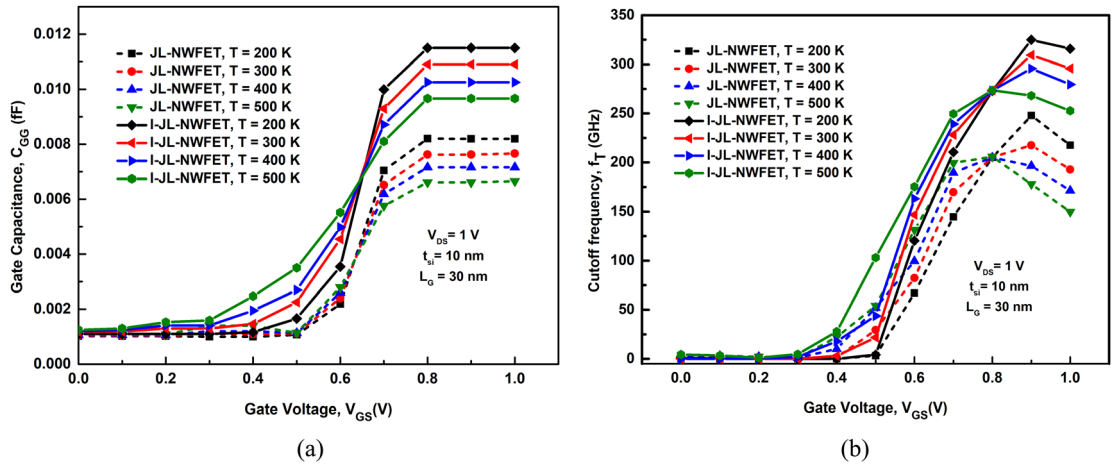


Figure 23. Plots of (a) gate capacitance and (b) cut-off frequency of I-JL-NWFET and JL-NWFET for different temperatures.

Table 5. Effect of temperature on the analog/RF performance of JL-NWFET and I-JL-NWFET.

Parameters	Devices	200 K	300 K	400 K	500 K
g_m (μS)	JL-NWFET	15.92	14.53	13.04	11.67
	I-JL-NWFET	23.43	21.21	19.53	17.70
TGF (V^{-1})	JL-NWFET	57.69	51.69	45.76	38.41
	I-JL-NWFET	84.91	80.98	76.61	72.75
g_d (μS)	JL-NWFET	1.16	1.27	1.56	1.95
	I-JL-NWFET	0.35	0.47	0.59	0.67
V_{EA} (V)	JL-NWFET	23.68	18.65	15.36	12.14
	I-JL-NWFET	55.63	49.14	39.14	33.94
C_{GG} (fF)	JL-NWFET	0.008	0.0070	0.0071	0.006
	I-JL-NWFET	0.011	0.0109	0.0102	0.009
f_T (GHz)	JL-NWFET	217.56	192.75	171.55	149.84
	I-JL-NWFET	315.80	295.80	279.54	252.73

at any given temperature, attributed to the increased outer fringe capacitance resulting from the utilization of the high-k spacer [17]. The behavior of f_T mirrors the trends observed in g_m and C_{GG}. It increases with a temperature rise at lower V_{GS} until a specific V_{GS}, after which it declines with further temperature increases, as demonstrated in figure 23(b). For the I-JL-NWFET, the decline in C_{GG} at higher V_{GS} is offset by an improvement in g_m, resulting in a superior f_T compared to the JL-NWFET. Notably, the reduction in f_T is 19.9% in I-JL-NWFET and 45.2% in JL-NWFET when the temperature is elevated from 200 K to 500 K.

The analog/RF parameters of JL-NWFET and I-JL-NWFET at different temperature are summarized in table 5. The analysis reveals that g_m, TGF, g_d, V_{EA}, C_{GG}, and f_T are less affected by temperature variations in I-JL-NWFET compared to JL-NWFET.

4.3. Impact of temperature on linearity parameters

In addition to achieving high speed, contemporary communication systems must ensure minimal signal distortion to assess the device's suitability for analog and RF applications. If a device fails to maintain linearity, the nonlinear components at the output may overwhelm the desired signal. To enhance the linearity of a system, the g_m must remain constant across the input voltage. In this subsection, the effect of temperature on the linearity performance of I-JL-NWFET is analyzed through metrics such as g_{m2}, g_{m3}, VIP2, VIP3, IMD3, and IIP3, and results are compared with JL-NWFET.

The effect of temperature on g_{m2} and g_{m3} against V_{GS} of I-JL-NWFET and JL-NWFET are portrayed in figures 24(a) and (b), respectively. The g_{m2} decreases as temperature increases in case of both I-JL-NWFET and JL-NWFET as exhibited in figure 24(a). Additionally, the peak value of g_{m2} of I-JL-NWFET is reduced by 2.53%, 5.45%, 10.74% and 11.67% as compared to JL-NWFET at 200 K, 300 K, 400 K and 500 K, respectively, indicating I-JL-NWFET is more linear than JL-NWFET at higher temperature. The linearity of the device is also associated with the flatness of the g_{m3} amplitude. The g_{m3} primarily contributes to the distortion of fundamental amplitude through signals in adjacent bands or intermodulation distortions. Moreover, the Zero-Crossover Point (ZCP) of

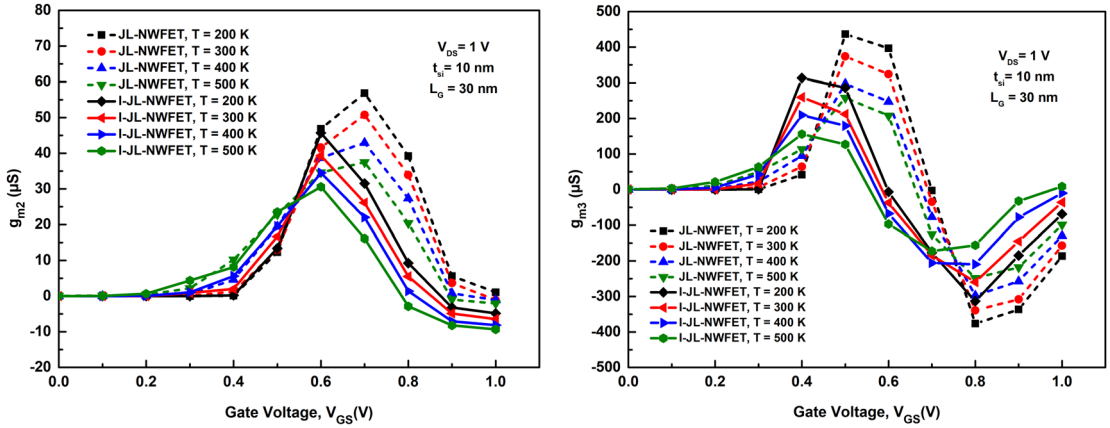


Figure 24. Plots of (a) g_{m2} and (b) g_{m3} of I-JL-NWFET and JL-NWFET for different temperatures.

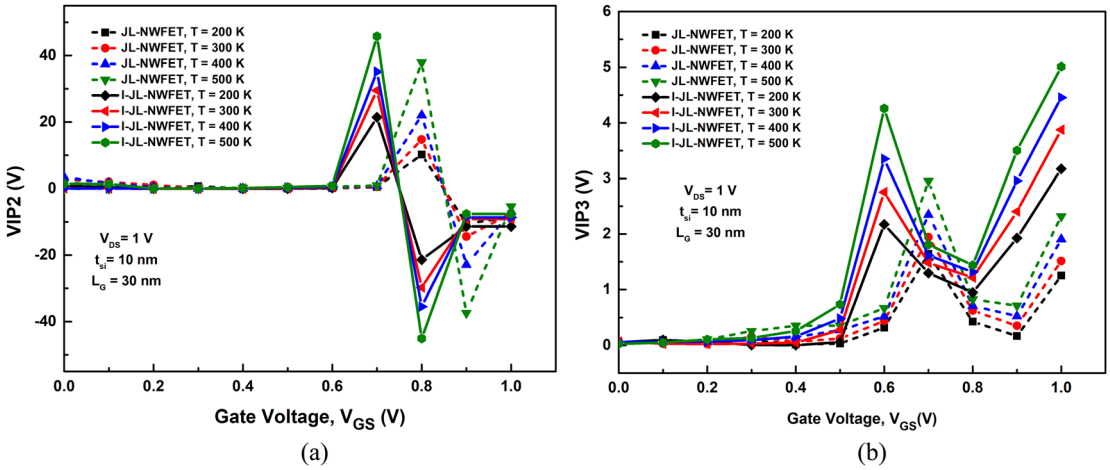


Figure 25. Plots of (a) VIP2 and (b) VIP3 of I-JL-NWFET and JL-NWFET for different temperatures.

g_{m3} should occur at a lower V_{GS} for improved linearity performance [20]. The V_{GS} value at which g_{m3} first crosses the zero axis is defined as ZCP and determines the DC bias point for optimal device operation. Furthermore, in I-JL-NWFET, ZCP is shifted towards lower V_{GS} compared to JL-NWFET at any given temperature. This shift is advantageous for achieving better linearity and lower distortion in device performance.

The influence of temperature on VIP2 and VIP3 of I-JL-NWFET and JL-NWFET are displayed in figures 25(a) and (b), respectively. It is observed in figure 25(a) that the peak value of VIP2 in I-JL-NWFET exhibits an increase by 2.09 times, 2.0 times, 1.59 times, and 1.20 times compared to that in JL-NWFET as the temperature escalates from 200 K to 500 K.

This augmentation is attributed to a reduction in the amplitude of g_{m2} with the rise of temperature. Furthermore, in figure 25(b), the peak value of VIP3 in I-JL-NWFET occurs when g_{m3} is zero and shows an increase by 32.3%, 41.7%, 43.1%, and 44% compared to that in JL-NWFET as the temperature increases from 200 K to 500 K.

Figures 26(a) and (b) depict the influence of temperature variation on IIP3 and IMD3 of I-JL-NWFET and JL-NWFET as functions of V_{GS}, respectively. The higher IIP3 peak observed in figure 26(a) of I-JL-NWFET than of JL-NWFET with rising temperature distinctly indicates enhanced linearity performance as the temperature increases. Additionally, IMD3 of I-JL-NWFET is found to be lower than that of JL-NWFET as depicted in figure 26(b). This observation ensures that in the I-JL-NWFET, HCEs have been mitigated, leading to a reduction in both power consumption and distortion of the device.

Table 6 summarized the less temperature variation of linearity parameter of I-JL-NWFET than that of JL-NWFET. Eventually, for increasing temperature (200 K - 500 K), I-JL-NWFET is reliable and useful in RF applications than JL-NWFET. Moreover, investigating alternative semiconductor materials with superior thermal stability could mitigate performance degradation at elevated temperatures. Material like silicon carbide (SiC) has

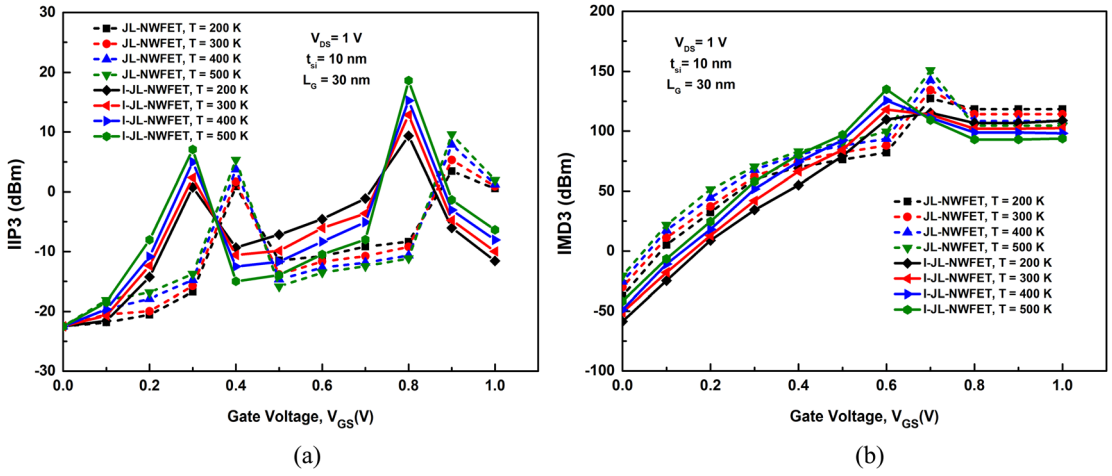


Figure 26. Plots of (a) IIP3 and (b) IMD3 of I-JL-NWFET and JL-NWFET for different temperatures.

Table 6. Effect of temperature on the linearity performance of JL-NWFET and I-JL-NWFET.

Parameters	Devices	200 K	300 K	400 K	500 K
Peak g_{m2} (μS)	JL-NWFET	56.82	50.75	42.89	37.47
	I-JL-NWFET	45.69	39.34	34.51	30.57
Peak g_{m3} (μS)	JL-NWFET	436.46	373.78	297.22	258.06
	I-JL-NWFET	313.35	259.73	209.30	155.98
Peak VIP2 (V)	JL-NWFET	10.22	14.74	22.05	37.93
	I-JL-NWFET	21.45	29.54	35.15	45.75
Peak VIP3 (V)	JL-NWFET	1.65	1.95	2.35	2.96
	I-JL-NWFET	2.18	2.76	3.35	4.26
Peak IIP3 (dBm)	JL-NWFET	3.42	5.31	7.87	9.60
	I-JL-NWFET	9.34	12.85	15.27	18.62
Peak IMD3 (dBm)	JL-NWFET	127.29	134.39	142.60	150.88
	I-JL-NWFET	109.65	118.04	125.63	134.96

exhibited notable potential in preserving electrical characteristics under at high temperatures [40]. It may serve as a potential future substitute for silicon material at high temperatures. Since this work has provided valuable insights through simulation analysis, forwarding to an experimental study may encounter challenges in fabricating a DMG in nanometer-scale technology. The fabrication process involves intricate steps such as contacting two gate materials, which introduces additional lithography complexity. Techniques like tilt angle deposition for different metal depositions further compound these challenges [41]. In response to this, the novel work function modulation method may be used which involves Molybdenum (Mo) as the gate material, allowing its work function to be adjusted by varying the nitrogen implantation in Mo [42]. This approach not only simplifies process integration but also opens avenues for enhancing device performance and scalability in future research efforts.

5. Conclusion

In this paper, a thorough investigation into the scaling of channel length and thickness for I-JL-NWFET is conducted and then the findings are compared with those of conventional JL-NWFET. The results demonstrate that the incorporation of an oppositely doped core, DMG, and high-k spacer in I-JL-NWFET leads to reduced sensitivity to the scaling of channel length and thickness compared to JL-NWFET. Furthermore, the effect of temperature on electrical, analog/RF, and linearity performance has been investigated for I-JL-NWFET and JL-NWFET. Notably, the electrical performance of I-JL-NWFET exhibits less variability than that of JL-NWFET. I_D experiences a reduction of 29% in I-JL-NWFET and 31.8% in JL-NWFET as the temperature changes from 200 K to 500 K. Moreover, variation in SS and DIBL of I-JL-NWFET are 29% and 28% less than in JL-NWFET. Analyzing the analog/RF performance of both I-JL-NWFET and JL-NWFET concerning parameters such as g_m, TGF, g_d, V_{EA}, C_{GG} and f_T based on temperature variation, it is observed that these parameter decreases with increase in temperature (200 K to 500 K). Besides this, due to rise in temperature from 200 K to 500 K, variation

of 24.4% (26.7%), 14.32% (33.40%), 32.9% (52.72%), and 19.9% (45.2%) in terms of g_m, TGF, V_{EA} and f_T for I-JL-NWFET (JL-NWFET) are observed. This indicates that I-JL-NWFET is less sensitive to temperature variations compared to JL-NWFET. Furthermore, the linearity performance of I-JL-NWFET, characterized by parameters such as g_{m2}, g_{m3}, VIP2, VIP3, IIP3 and IMD3 is improved at higher temperature than that of JL-NWFET. Therefore, temperature emerges as a significant factor affecting device characteristics, where an increase in temperature suppresses electrical and analog/RF performance while enhancing the linearity performance of I-JL-NWFET when compared to JL-NWFET.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Ethics approval

The manuscript is not submitted to anywhere else for the publication. The submitted work is an original work and have not been published elsewhere in any form or language. The authors do not have any conflict of interest. The research does not involve any human participant and/or animal thus, informed consent in not required.

Consent for publication

All the authors understand and give their consent for the publication of manuscript 'Investigating the Effect of Scaling and Temperature on the Performance of Improved Junctionless Nanowire FET through simulation Analysis' in 'Physica Scripta' Journal.

Competing interests

No, all the authors declare that they have no competing interests as defined by IOP Science, or other interests that might be perceived to influence the results and/or discussion reported in this paper. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Funding

Not applicable

Author's contributions

Bharti: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing—original draft.
Poornima Mittal: Conceptualization, Investigation, Methodology, Supervision, Visualization, Writing—review & editing.

ORCID iDs

Poornima Mittal https://orcid.org/0000-0002-9479-8628

References

[1] Oh S-H, Monroe D and Hergenrother J M 2000 Analytic description of short-channel effects in fully depleted double-gate and cylindrical, surrounding-gate MOSFETs *IEEE Electron Device Lett.* **21** 445–7
[2] Song J, Yu B, Yuan Y and Taur Y 2009 A review on compact modeling of multiple-gate MOSFETs *IEEE Trans. Circuits Syst. I Regul. Pap.* **56** 1858–69
[3] Lee C-W, Afzalian A, Akhavan N D, Yan R, Ferain I and Colinge J-P 2009 Junctionless multigate field-effect transistor *Appl. Phys. Lett.* **94** 053511
[4] Gupta S, Mittal P and Juneja P 2022 Performance Improvement in single-gate organic transistors with contacts at top and bottom: Additional p + region insertion near source and drain *Main Group Chem.* **21** 243–61
[5] Negi S, Mittal P and Kumar B 2020 Numerical modeling and parameters extraction of novel triple hole block layer-based organic light-emitting diode for display *J. Soc. Inf. Disp.* **28** 956–64

[6] Maurya P, Mittal P and Kumar B 2022 Performance improvement for organic light emitting diodes by changing the position of mixed-interlayer *Main Group Chem.* **21** 837–49
[7] Gupta S, Mittal P and Juneja P 2023 Ditch incorporated organic thin film transistor based organic all-p inverter: a novel approach *Phys. Scr.* **98** 115038
[8] Chamola P and Mittal P 2023 PPV–PCBM bulk heterojunction organic solar cell to power modern pacemakers *J. Mater. Res.* **38** 1304–16
[9] Mittal P, Yadav S and Negi S 2021 Advancements for organic thin film transistors: Structures, materials, performance parameters, influencing factors, models, fabrication, reliability and applications *Mater. Sci. Semicond.* **133** 105975
[10] Colinge J P *et al* 2010 Nanowire transistors without junctions *Nat. Nanotechnol.* **5** 225–9
[11] Sahay S and Kumar M J 2016 Controlling L-BTBT and volume depletion in nanowire JLFETs using core–shell architecture *IEEE Trans. Electron Dev.* **63** 3790–4
[12] Bharti and Mittal P 2024 Oppositely-doped core–shell junctionless nanowire FET: Design and investigation *ECS J. Solid State Sci. Technol.* **13** 013004
[13] Lou H, Li D, Dong Y, Lin X, He J, Yang S and Chan M 2013 Suppression of tunneling leakage current in junctionless nanowire transistors *Semicond. Sci. Technol.* **28** 125016
[14] Lou H, Zhang L, Zhu Y, Lin X, Yang S, He J and Chan M 2012 A junctionless nanowire transistor with a dual-material gate *IEEE Trans. Electron Dev.* **59** 1829–36
[15] Xie H, Liu H, Wang S, Chen S, Han T and Li W 2019 Improvement of electrical performance in heterostructure junctionless TFET based on dual material gate *Appl. Sci.* **10** 126
[16] Gundapaneni S, Ganguly S and Kottantharayil A 2011 Enhanced electrostatic integrity of short-channel junctionless transistor with high-k spacers *IEEE Electron Device Lett.* **32** 1325–7
[17] Baruah R K and Paily R P 2014 A dual-material gate junctionless transistor with high-k spacer for enhanced analog performance *IEEE Trans. Electron Dev.* **61** 123–8
[18] Hammoud A, Patterson R L, Gerber S and Elbuluk M 2003 Electronic components and circuits for extreme temperature environments *10th IEEE International Conference on Electronics, Circuits and Systems* 1, 44–7
[19] Ma W, Kaya S and Asenov A 2003 Study of RF linearity in sub-50 nm MOSFETs using simulations *J. Comput. Electron.* **2** 347–52
[20] Ghosh P, Haldar S, Gupta R S and Gupta M 2012 An investigation of linearity performance and intermodulation distortion of GME CGT MOSFET for RFIC design *IEEE Trans. Electron Dev.* **59** 3263–8
[21] Pratap Y, Haldar S, Gupta R S and Gupta M 2014 Performance evaluation and reliability issues of junctionless CSG MOSFET for RFIC design *IEEE Trans. Device Mater. Reliab.* **14** 418–25
[22] Zhang L *et al* 2016 Reliability and performance evaluation for STT-MRAM under temperature variation *2016 17th Int. Conf. on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems (EuroSimE), Montpellier, France* 1–4
[23] Kalambe J and Patrikar R 2014 Design and development of microcantilever based detector *2014 Int. Conf. on Electronic Systems, Signal Processing and Computing Technologies, Nagpur, India* 335–40
[24] Kumar N, Awasthi H, Purwar V, Gupta A and Dubey S 2021 Impact of temperature variation on analog, hot-carrier injection and linearity parameters of nanotube junctionless double-gate-all-around (NJL-DGAA) MOSFETs *Silicon* **14** 2679–86
[25] Gupta A, Gupta V, Pandey A K and Gupta T K 2022 A novel technique to investigate the impact of temperature and process parameters on electrostatic and analog/RF performance of channel modulated junctionless gate-all-around (CM-JL-GAA) MOSFET *Silicon* **14** 10613–22
[26] Atlas Device Simulation Software Silvaco Int. Santa Clara. USA, CA 2018
[27] Choi S J, Moon D I, Kim S, Duarte J P and Choi Y K 2011 Sensitivity of threshold voltage to nanowire width variation in junctionless transistors *IEEE Electron Device Lett.* **32** 125–7
[28] Fan J, Li M, Xu X, Yang Y, Xuan H and Huang R 2015 Insight into gate-induced drain leakage in silicon nanowire transistors *IEEE Trans. Electron Dev.* **62** 213–9
[29] Trevisoli R *et al* 2016 Junctionless nanowire transistors operation at temperatures down to 4.2 K *Semicond. Sci. Technol.* **31** 114001
[30] Liu Y, Guo J, Zhu E, Liao L, Lee S, Ding M, Shakir I, Gambin V, Huang Y and Duan X 2018 Approaching the Schottky–Mott limit in van der Waals metal–semiconductor junctions *Nature* **557** 696–700
[31] Gundapaneni S, Bajaj M, Pandey R K, Murali K V R, Ganguly S and Kottantharayil A 2012 Effect of band-to-band tunneling on junctionless transistors *IEEE Trans. Electron Dev.* **59** 1023–9
[32] Mittal P, Kumar B, Negi Y S, Kaushik B K and Singh R K 2012 Channel length variation effect on performance parameters of organic field effect transistors *Microelectron. J.* **43** 985–94
[33] Wang Y, Shan C, Dou Z, Wang L and Cao F 2015 Improved performance of nanoscale junctionless transistor based on gate engineering approach *Microelectron. Reliab.* **55** 318–25
[34] Sahay S and Kumar M J 2017 Diameter dependence of leakage current in nanowire junctionless field effect transistors *IEEE Trans. Electron Dev.* **64** 1330–5
[35] Mittal P 2021 Ditch and elevated organic thin film transistor-based improved common source voltage amplifier: Frequency response characteristics and analytical modelling *J. Soc. Inf. Disp.* **29** 642–58
[36] Kumar A, Gupta T K, Shrivastava B P and Gupta A 2023 Impact of temperature variation on noise parameters and HCI degradation of Recessed Source/Drain Junctionless Gate All Around MOSFETs *Microelectron. J.* **134** 105720
[37] Awasthi H *et al* 2021 Impact of temperature on analog/RF performance of dielectric pocket gate-all-around (DPGAA) MOSFETs *Silicon* **13** 2071–5
[38] Han M-H, Chen H-B, Yen S-S, Shao C-S and Chang C-Y 2013 Temperature-dependent characteristics of junctionless bulk transistor *Appl. Phys. Lett.* **103** 133503
[39] Datta E, Chattopadhyay A and Mallik A 2020 Relative study of analog performance, linearity, and harmonic distortion between junctionless and conventional SOI FinFETs at elevated temperatures *J. Electron. Mater.* **49** 3309–16
[40] Lu X, Wang L, Yang Q, Yang F, Gan Y and Zhang H 2023 Investigation and comparison of temperature-sensitive electrical parameters of SiC mosfet at extremely high temperatures *IEEE Trans. Power Electron.* **38** 9660–72
[41] Jain A K, Sahay S and Kumar M J 2018 Controlling L-BTBT in emerging nanotube FETs using dual-material gate *IEEE J. Electron Devices Soc.* **6** 611–21
[42] Li T-L, Hu C-H, Ho W-L, Wang H C-H and Chang C-Y 2005 Continuous and precise work function adjustment for integratable dual metal gate CMOS technology using Hf-Mo binary alloys *IEEE Trans. Electron Dev.* **52** 1172–9

Iris and Palmprint Multimodal Biometric Recognition using Novel Preactivated Inverted ResNet and Hybrid Metaheuristic Optimized DenseNet

Indu Singh[1], Shruti Khatri[2], Sunaina Luthra[3], Tanvi Singh[4], Gunbir Singh Baveja[5],
[1,2,3,4]Department of Computer Science, Delhi Technological University, Delhi
[5]Department of Computer Science, University of British Columbia, Vancouver

arXiv:2407.03498v1 [cs.CV] 3 Jul 2024

Biometric recognition technology has witnessed widespread integration into daily life due to the growing emphasis on information security. In this domain, multimodal biometrics, which combines multiple biometric traits, has overcome limitations found in unimodal systems like susceptibility to spoof attacks or failure to adapt to changes over time. This paper proposes a novel multimodal biometric recognition system that utilizes deep learning algorithms using iris and palmprint modalities. A pioneering approach is introduced, beginning with the implementation of the novel Modified Firefly Algorithm with Lévy Flights (MFALF) to optimize the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm, thereby effectively enhancing image contrast. Subsequently, feature selection is carried out through a unique hybrid of ReliefF and Moth Flame Optimization (MFOR) to extract informative features. For classification, we employ a parallel approach, first introducing a novel Preactivated Inverted ResNet (PIR) architecture, and secondly, harnessing metaheuristics with hybrid of innovative Johnson Flower Pollination Algorithm and Rainfall Optimization Algorithm for fine tuning of the learning rate and dropout parameters of Transfer Learning based DenseNet architecture (JFPA-ROA). Finally, a score-level fusion strategy is implemented to combine the outputs of the two classifiers, providing a robust and accurate multimodal biometric recognition system. The system's performance is assessed based on accuracy, Detection Error Tradeoff (DET) Curve, Equal Error Rate (EER), and Total Training time. The proposed multimodal recognition architecture, tested across CASIA Palmprint, MMU, BMPD, and IIT datasets, achieves 100% recognition accuracy, outperforming unimodal iris and palmprint identification approaches.

***Index Terms*—Multimodal biometric Recognition, Metaheuristic Optimisation, Convolutional Neural Network, Feature Selection, Deep Learning, Transfer Learning**

I. Introduction

In today's digital landscape, traditional authentication systems relying on passwords and tokens are insufficient against modern cybersecurity threats like brute-force attacks and phishing exploits. These vulnerabilities leave sensitive data and systems at risk. The Biometric based recoginition system has emerged as a compelling solution to these shortcomings [1].

Biometrics involves identifying individuals based on their physiological or behavioral traits. This technology includes a diverse range of characteristics, from behavioral traits like signature dynamics, voice patterns, and gait analysis, to physiological traits such as fingerprints, iris patterns, and palmprints [2]. Biometric recognition systems automate this identification process, representing a fundamental shift in authentication methods.

Unimodal biometric systems rely on a single characteristic like iris patterrns or palmprints, for identification. These traits are particularly significant due to their permanence, uniqueness, and high resistance to forgery, which ensures high recognition accuracy and valuable security [3, 4]. However, unimodal systems face limitations. Factors like injuries, sensor noise, or temporary changes in appearance can affect their accuracy. Multimodal biometric authentication addresses these issues by combining multiple biometric modalities, thereby enhancing security against imposters who would need to replicate several unique identifiers [5].

Machine learning further enhances multimodal biometric systems by improving recognition accuracy and adapting to new data. This technology addresses the limitations of unimodal systems, making biometric recognition more reliable and secure.

Traditionally, biometrics researchers relied on machine learning algorithms for identification, which involved multiple steps: data cleaning, feature extraction, and classification. However, these methods have limitations, including unreliable feature extraction and difficulty with common image transformations. To overcome these limitations, many researchers have recently started using deep learning-based biometric systems. Deep learning significantly improves performance by automatically extracting features directly from raw biometric data, streamlining the process and enhancing robustness [6]. However, optimizing the fusion of multimodal features for accurate recognition remains complex. Metaheuristic algorithms play a crucial role in this context due to their efficiency in navigating complex search spaces and identifying optimal solutions. They offer a robust framework for effectively integrating multiple biometric modalities [7]. Leveraging sophisticated algorithms and pattern recognition techniques, biometric systems provide strong authentication by capturing and analyzing distinctive traits, ensuring secure access to sensitive information and safeguarding against unauthorized intrusion [8].

However, despite the advancements in biometric integration, a notable gap exists in the literature regarding the fusion of iris and palmprint recognition using deep learning methodologies. While metaheuristic algorithms offer promise in addressing these challenges, the direct fusion of iris and palmprint modal-

ities remains limited. Moreover, existing multimodal biometric systems employing deep learning often encounter limitations, including inadequate image preprocessing, feature selection, and classification techniques, which impede their effectiveness. Additionally, challenges such as rigid classification approaches susceptible to overfitting and scalability issues emphasize the critical need to bridge the gap between metaheuristic algorithms and deep learning methodologies to develop more robust and effective biometric recognition systems. [9, 10, 11].

This paper proposes a comprehensive approach to address the aforementioned limitations by integrating iris and palmprint modalities using deep learning and metaheuristic algorithms. Firstly, we enhance image contrast effectively by utilizing the Modified Firefly Algorithm with Lévy Flights (MFALF) to optimize the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm. This optimization significantly improves the quality of input data. Secondly, we employ a unique hybrid of ReliefF and Moth Flame Optimization (MFOR) for feature selection. This approach ensures the extraction of informative features, enhancing the discriminative power of our system.

For classification, we adopt a parallel approach. Firstly, we introduce the innovative Preactivated Inverted ResNet (PIR) architecture to enhance classification performance. Secondly, we utilize metaheuristic algorithms, specifically a hybrid of the Johnson Flower Pollination Algorithm and Rainfall Optimization Algorithm, to fine-tune the learning rate and dropout parameters of the Transfer Learning-based DenseNet architecture (JFPA-ROA-DenseNet). This fine-tuning process optimizes model parameters, mitigating overfitting and improving generalization capability. Finally, we implement a score-level fusion strategy to combine the outputs of the two classifiers. This strategy ensures the development of a robust and accurate multimodal biometric recognition system by leveraging the strengths of both classifiers.

The key findings of this research are outlined below:

1) *We present a novel combination of Modified Firefly Algorithm with Lévy Flights (MFALF) to optimize the CLAHE algorithm for image contrast enhancement.*
2) *For enhancing feature selection, we introduce a unique hybrid of ReliefF and Moth Flame Optimization to extract informative features.*
3) *We present a parallel classification model encompassing transfer learning based novel Preactivated Inverted Resnet (PIR) architecture and an innovative hybrid metaheuristic optimized DenseNet architecture.*
4) *An innovative hybrid of Johnson Flower Pollination Algorithm and Rainfall Optimization Algorithm is proposed for fine-tuning parameters of Transfer Learning based DenseNet (TL-JFPA-ROA-DenseNet) for preventing overfitting.*

The rest of this paper is organized as follows: Section II provides an overview of previous studies. Section III presents the dataset description. Section IV outlines the proposed methodology, including preprocessing, classification, and score level fusion. Section V discusses the experimental setup and Implementation, while Section VI reports the obtained results. Section VII summarizes key findings and proposes future research directions, concluding the paper.

II. Literature Review

Numerous studies have presented multimodal biometric systems employing diverse recognition techniques. In this section, we review the existing literature related to multimodal biometric recognition, metaheuristic algorithms, and deep learning models.

A. Image Contrast Enhancement using CLAHE

Image enhancement plays a crucial role in various image processing applications. It improves the visual quality of images, making them easier to analyze for humans and facilitating further image processing tasks.

Image enhancement often faces challenges with uneven illumination, prompting the use of histogram equalization (HE). Yet, Salem et al. (2019) [12] noted limitations and explored alternatives. Their study evaluated HE alongside Contrast-Limited Adaptive Histogram Equalization (CLAHE), Cumulative Histogram Equalization (CHE), and Quadrant Dynamic Histogram Equalization (QDHE). They found CLAHE to outperform others, particularly in handling non-uniform intensity distribution.

CLAHE's success extends to diverse fields, notably biometric recognition, with studies by Nour et al. (2024) [13], Vidyarthi and Malik (2022) [14], and Lv et al. (2020) [15]. However, its manual parameter selection remains a time-consuming and inconsistent drawback.

Automatic parameter selection for CLAHE is an area of limited exploration in the literature. However Surya and Muthukumaravel (2023) [16] proposed an Adaptive Sailfish Optimization-Contrast Limited Adaptive Histogram Equalization (ASFO-CLAHE) method, which leverages an optimization algorithm to automatically determine the clip limit and number of tiles – key parameters in CLAHE. Similarly, Campos et al. (2019) [17] presented a supervised learning approach for parameter selection

These advancements in automatic parameter selection hold promise for enhancing the efficiency and effectiveness of CLAHE in various image processing applications.

B. Hybrid Feature Selection for Multimodal Biometric Recognition

Feature selection is pivotal in machine learning, targeting the identification of pertinent features to optimize model efficiency and performance. Agrawal et al. (2021) [18] offer an extensive review of feature selection employing metaheuristic algorithms from 2009 to 2019, categorizing approaches into filter, wrapper, and embedded methods. Filter methods assess feature relevance independently of the learning algorithm like information gain and ReliefF, while wrapper methods involve evaluating subsets of features based on model performance using metaheuristic algorithms like Moth-Flame Algorithm (MFA) and Genetic Algorithm (GA). Embedded methods integrate feature selection directly into the training process using regularization techniques.

Researchers have investigated diverse feature selection methods, like Trabelsi et al. (2022) [19] who utilized filter methods Fisher score and ReliefF algorithms. Gona et al. (2023) [20] introduced a wrapper method bio-optimization-based modified Lion optimization algorithm (MLOA) for optimizing feature selection by uncovering dependencies among biometrics.

Hsu et al. (2011) [21] suggested using a single filter or wrapper method for feature selection can lead to suboptimal results, as filters may miss feature interactions and wrappers are computationally expensive. Thus, a hybrid approach is needed to balance efficiency and accuracy. Dabba et al. (2021) [22] created a hybrid by merging Mutual Information Maximization and the Moth Flame Algorithm to assess gene relevance and redundancy, evolving subsets based on fitness functions. Similarly, Mafarja and Mirjalili (2017) [23] utilized a hybrid strategy combining Whale Optimization Algorithm with simulated annealing for feature selection.

These hybrid methods leverage the strengths of different approaches to achieve superior performance and accuracy in feature selection tasks.

C. Deep Learning for Biometric recognition

Deep learning models have revolutionized biometrics by providing advanced techniques for pattern recognition and classification tasks, resulting in high accuracy and robust performance.

Chollet (2018) [24] highlighted the reluctance to train deep learning models from scratch due to the time and extensive datasets involved. Instead, researchers prefer pre-trained models like Inception, VGGNet, AlexNet, and DenseNet, enabling faster development of deep learning-based biometric applications through fine-tuning and transfer learning.

Nguyen et al. (2017) [25] compared the effectiveness of five CNN pre-trained models—VGGNet, Inception, AlexNet, ResNet, and DenseNet—for iris recognition, with DenseNet yielding the highest accuracy. Meanwhile, Wang and Kumar (2019) [26] proposed an innovative approach utilizing ResNet with dilated convolutional kernels, resulting in substantial accuracy enhancements in iris recognition compared to traditional methods.

Zhang et al.(2018) [27] proposed a DCNN-based method for contactless palmprint and palmvein recognition using modified Inception_ResNet_v1 architecture for feature extraction, paired with an SVM classifier for identification.

While pre-trained models offer a good starting point, neglecting hyperparameter tuning can hinder the performance. Tuba et al. (2021) [28] emphasize its significance, particularly in adjusting the learning rate and implementing techniques to mitigate overfitting. Rosa et al. (2018) [29] explored dropout with meta-heuristic optimization to address overfitting in CNNs for biometric recognition.

Hence, Deep learning revolutionizes biometrics with precise recognition. However, selecting pre-trained models, new architectures, and hyperparameter tuning are critical for optimal performance.

D. Score level Fusion for multimodal Biometric Recognition

Lip and Ramli (2012) [30] highlight the significance of fusion strategies in multimodal biometrics, stressing the need to combine various modalities for improved authentication accuracy. They particularly emphasize three fusion levels: feature, score, and decision. Feature-level fusion combines features, score-level fusion merges model scores, and decision-level fusion integrates decisions from multiple models. Score-level fusion is preferred for its simplicity and effectiveness in distinguishing genuine from impostor scores.

Score-level fusion, noted for its simplicity and efficiency, has been extensively researched. Noh et al. (2020) [31] applied it to output scores from deep CNNs for finger-vein recognition, while He et al. (2010) [32] investigated sum rule and SVM-based fusion for fingerprint, face, and finger vein data. Additionally, Peng et al. (2014) [33] proposed a multimodal authentication approach combining various biometric scores using t-norms, offering reduced computational complexity without the need for learning or training.

Such fusion strategies enhance biometric system performance by leveraging the strengths of different modalities and reducing vulnerabilities inherent in single-modality approaches.

III. DATASET DESCRIPTION

The following section details the data collection process, describing the sources, methods, and protocols employed to gather the datasets used in this study.

Two distinct biometric traits, palmprint and iris, were utilized in our experiments.We have used MMU1 dataset [34] for collecting iris image & three datasets: CASIA-palmprint V1 (CASIA) [35], IIT Delhi Touchless Palmprint (IIT) [36], and Birjand University Mobile Palmprint (BIRJ) [37] were used for collecting palmprint images.The above mentioned unimodal datasets were selected due to their public availability, extensive use in numerous biometric research studies, and well-established status in the field.

A. Unimodal datasets

In the following points, we describe each unimodal dataset, separately.

- The MMU1[34] publicly avaiable iris dataset comprises 450 images, with 5 images each of left and right iris of 45 subjects. Training involves 8 images per subject, and testing uses 2 images per subject, maintaining an 80:20 train-to-test ratio.
- In the CASIA[35] palmprint dataset, there are 5,502 palmprint images of 312 subjects,and each palmprint image is an 8 bit gray-level JPEG file. To accommodate system limitations, 45 subjects were randomly chosen for evaluating our proposed methodology. The model is trained on both left and right palm print images, with a train-to-test ratio of 80:20 .
- The IIT[36] dataset encompasses 2,760 left and right palmprint images from 230 subjects acquired with a touchless setup, providing images in bitmap format with

variations in hand pose and scale. Our model has been trained over 1,380 left and right palmprint images of all 230 subjects. Testing involves 460 images, maintaining a train-to-test ratio of 75:25 . It is important to note that The PolyU palmprint database is widely used but relies on user-pegs to control variations in hand-pose and image scale, although inconvenient for users. This standardization is a key factor contributing to the notably high recognition accuracy, nearly 100%, achieved in version 2.0 of the database. In contrast, IIT Delhi's touchless palmprint database, aims for a user-friendly and reliable palmprint identification using completely touchless imaging setup.

- Regarding the Birjand[37] palm print images, the dataset comprises 1,640 images representing both left and right palm prints from 41 subjects, taken in two sessions two weeks apart. In the first session, six images of each palm were captured from a 20 cm distance, with no strict hand orientation control. In the second session, 16 images per palm were taken with increased hand rotation, promoting intraclass variability by having users close and open their hands between captures. Training and evaluation of the proposed model for all 41 subjects are conducted with an 80:20 train-to-test ratio.

B. Multimodal datasets

Due to the unavailability of multimodal biometric datasets, we have created three multimodal datasets for biometric evaluation:

- M-CAS: This dataset has been created by combining iris images from 45 subjects in the MMU Iris dataset with palmprint images from the same 45 subjects in the CASIA palmprint dataset, previously chosen for unimodal evaluation
- M-IIT: integrates palmprint and iris images by pairing 45 subjects from the MMU Iris dataset with randomly selected 45 subjects from the test dataset of the IIT palmprint dataset, which was originally used in Unimodal biometric recognition evaluation.
- M-BIRJ: Following a similar methodology as M-IIT, M-BIRJ dataset was developed comprising iris and palmprint images of 45 subjects.

The novel multimodal datasets discussed above enable comprehensive evaluation of multimodal biometric recognition systems, providing researchers and developers with valuable resources for testing and advancing biometric authentication technologies. Detailed description of our proposed approach is discussed in subsequent section.

IV. Methodology

In this section, we outline our proposed method for identifying individuals through a multimodal biometric recognition system utilizing both Iris and Palmprint data. The proposed methodology initiates by preprocessing the datasets. We aim to enhance image quality through Contrast Limited Adaptive Histogram Equalization (CLAHE), optimized using the novel Modified Firefly Algorithm with Lévy Flights (MFALF) for noise reduction and improved local contrast. Following this, we employ a unique blend of ReliefF and Moth Flame Optimization (MFOR) for feature selection to extract informative traits. The extracted features are then fed for classification in parallel. We introduce the innovative Preactivated Inverted ResNet (PIR) architecture and fine-tune the Transfer Learning based DenseNet by harnessing metaheuristics with hybrid of innovative Johnson Flower Pollination Algorithm and Rainfall Optimization Algorithm (JFPA-ROA-DenseNet). Finally, we merge classifier outputs through a score-level fusion strategy to deliver a robust and accurate multimodal biometric recognition system. The proposed model is illustrated in Fig. 1.

A. Contrast Limited Adaptive Histogram Equalization - Modified Firefly Algorithm with Lévy Flights (CLAHE-MFALF)

Inspired by Surya and Muthukumaravel (2023) [16], the proposed methodology exploits CLAHE preprocessing technique by optimizing it's clip limit parameter by novel MFALF algorithm. In CLAHE [38] the clip limit parameter sets the threshold for limiting the contrast enhancement in local image regions.

MFALF Algorithm

The Firefly Algorithm is a metaheuristic optimization technique inspired by the flashing behavior of fireflies. Firefly algorithm utilizes the behavior of fireflies in a search space, with attractiveness determined by fitness values. Brighter fireflies signal higher attractiveness, and through iterations, fireflies move towards brighter peers, influenced by distance and randomization.

For computing fitness function and attractiveness value of each "firefly" parameters like PSNR (Peak Signal-to-Noise Ratio), SSIM (Structural Similarity Index) are used emphasizing on quality and structural similarity of images. PSNR is the ratio of the maximum possible signal power to the power of the noise that affects the image. It quantifies image quality by measuring noise in optimized images.

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{B_{Max}^2}{B_{MSE(i,j)}} \right), \tag{1}$$

Where, B_{Max} is the maximum brightness value in a target image (255 in an 8-bit grayscale image), B_{MSE} is the Mean Squared Error between the brightness values of the estimated (i) and high-resolution images (j).

SSIM is a metric that evaluates the structural similarity between the original and optimized images.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2}$$

Where x and y are the original and optimized image respectively, μ_x is the mean of x, σ_x^2 is the variance of x, and σ_{xy} is the covariance of x and y. C_1 and C_2 are constants for stabilization.

Definition 1. ***Modified Attractiveness Parameter*** ***(β')***
It is a tuning parameter that governs the strength of attraction between fireflies. It is updated based on euclidean distance r.

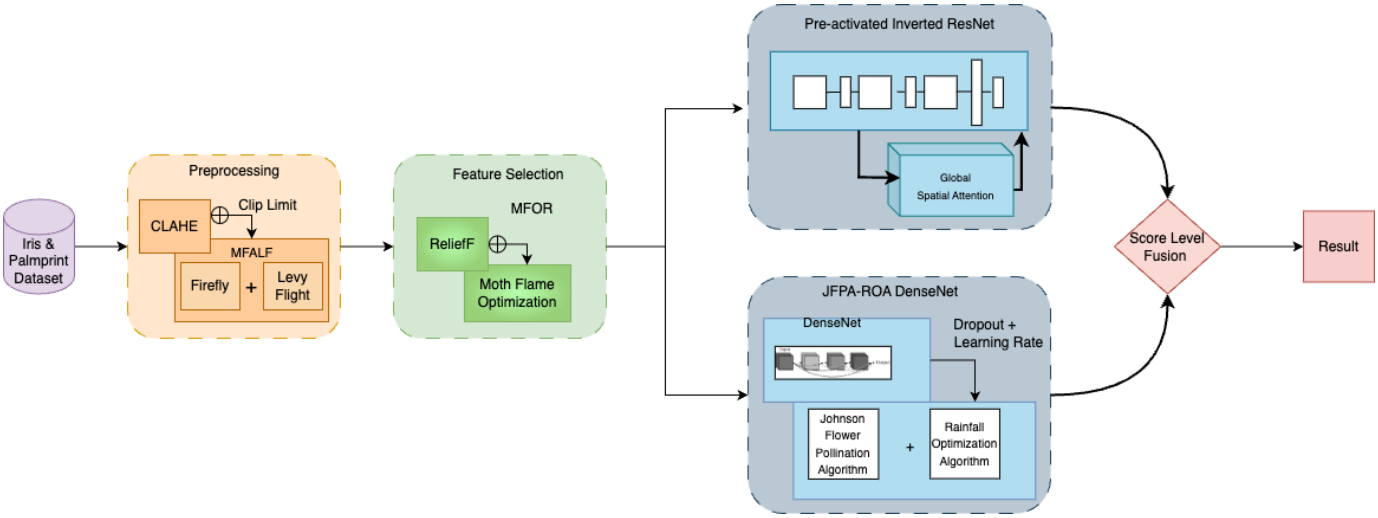


Fig. 1: Architecture of proposed model

$$r = \|\vec{X}_j - \vec{X}_i\| \tag{3}$$

$$\beta = \beta_0 \cdot e^{-\gamma \cdot \mathrm{r}^2} \tag{4}$$

Where, β_0 is a self tuned constant, γ is the absorption coefficient which controls the rate at which the β decreases and ranges between [0.1,1].

However this equation does not take into consideration the Weights of both the fire flies which is important according to the newton's law of attraction. Thus, We propose a modified equation for calculation of attraction :

$$\beta' = \beta_0 e^{\left(-\frac{\gamma r^2}{m_1 \cdot m_2}\right)} \tag{5}$$

Where m_1 and m_2 are decided by the fitness value of the firefly in population. We define the fitness function (F) and the firefly-attractiveness function (A) as :

$$F(\alpha) = \alpha \cdot \text{SSIM} + (1-\alpha) \cdot \text{PSNR} \tag{6}$$

$$A(i,j) = \beta e^{\left(-(F)^2\right)} \tag{7}$$

Here, α is a randomization parameter ranging [0,1] and β is attractiveness parameter ranging [0,1].

Definition 2. ***Levy Weighing Factor***

Levy Weighing Factor is a parameter which modulates the magnitude of random displacements in each iteration. It is proposed to introduce Levy flight, as an integral component of the Firefly Algorithm by modifying the equation 8 for new position of firefly as depicted in eq. 9.

$$x(i, i+1) = x(i,t) + A(i,j) \cdot F \tag{8}$$

$$x(i, i+1) = x(i,t) + A(i,j) \cdot F + W \cdot (rand() - 0.5) \tag{9}$$

Here, $x(i,j)$ denotes the firefly value at position (i,j) and W is the Levy Weighing Factor introduced in the model.

Levy flight's stochastic behavior reduces the likelihood of getting stuck in local optima and increases the algorithm's chances of finding a globally optimal solution.

Algorithm 1 describes the MFALF algorithm. Here, n is the population size, J is the search space, seed refers to the initial configuration of fireflies in the search space, $\alpha, \beta,$ andγ are the randomization, attractiveness and the absorption parameters respectively. Original and preprocessed images of palmprint and iris are shown in Fig 2 and Fig 3 respectively.

The preprocessed images are further passed in the model for extraction of relevant features such that dimensioality is reduced. The preprocessed images are then passed through the model to extract relevant features, effectively reducing dimensionality.

Algorithm 1 Modified Firefly Algorithm with Levy Flight

Require: $n, \alpha, \beta, \gamma,$ seed

Initialize fireflies $\vec{x}_1, \ldots, \vec{x}_n \in [\text{lb}, \text{ub}]$ with intensities $s_1, \ldots, s_n$ sampled from Population P
best $\leftarrow \min_i s_i$
$t \leftarrow n$
$\alpha' \leftarrow \alpha$
while $t \leq t_{\max}$ **do**
  $\alpha' \leftarrow \alpha' \cdot 0.97$
  for $i \in [1, n]$ **do**
    for $j \in [1, n]$ **do**
      if $s_i \geq s_j$ **then**
        $r \leftarrow \|\vec{x}_i - \vec{x}_j\|^2$
        $\beta \leftarrow \beta \cdot e^{-r\gamma}$
        Steps$\leftarrow \alpha' \cdot (\text{rand(dim)} - 0.5) \cdot (\text{ub} - \text{lb}))$
        $\vec{x}_i \leftarrow \vec{x}_i + \beta \cdot (\vec{x}_j - \vec{x}_i) + \text{steps}$
        $s_i \leftarrow J(\vec{x}_i)$
        $t \leftarrow t + 1$
        best $\leftarrow \min(s_i, \text{best})$
      end if
    end for
  end for
end while
return best

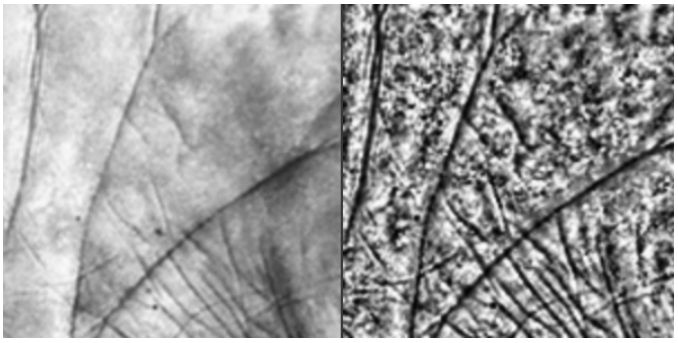


Fig. 2: Original and Preprocessed Palmprint Image

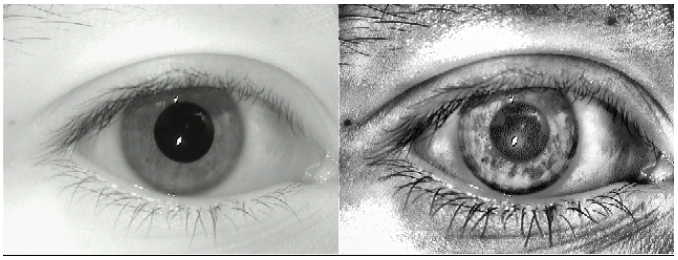


Fig. 3: Original and Preprocessed Iris Image

B. Moth Flame Optimization-ReliefF (MFOR)

Our proposed novel hybrid feature selection approach, termed MFOR, synergistically combines the robust feature ranking capabilities of filter method ReliefF with the efficient subset search strategy of Moth Flame Optimization (MFO) as a wrapper method. The hybrid can be categorized in 2 phases: *ReliefF* and *MFO*.

1) Phase I - ReliefF Phase

In the ReliefF phase, relevant features are selected based on their importance using ReliefF, and the corresponding subset is used for further optimization.

The original Relief algorithm [39] initially assigns equal weights to all features, then updates the feature scores ($W[f]$) for feature (f) as :

$$W[f] = \left(\frac{|x_i^{(A)} - x_i^{(M_j)}| - |x_i^{(A)} - x_i^{(H_j)}|}{\max(x_i) - \min(x_i)} \right) \tag{10}$$

Where $x_i^{(A)}$ is the value of the i-th feature for the instance being considered, $x_i^{(M_j)}$ and $x_i^{(H_j)}$ are the values of the i-th feature for the nearest instance of different ($Miss$) and same classes (Hit), respectively.

ReliefF [40] introduces key improvements: firstly, it incorporates a 'number of neighbors' parameter (k) for scoring updates, using k nearest hits and misses instead of just one, enhancing reliability, especially in noisy datasets. The ReliefF score for feature f is calculated as :

$$W[f] = \sum_{j=1}^{k} \frac{1}{k} \left(\frac{diff(x_i^{(A)}, x_i^{(M_j)})^2 - diff(x_i^{(A)}, x_i^{(H_j)})^2}{\max(x_i) - \min(x_i)} \right) \tag{11}$$

Where, k is the number of nearest neighbors and diff(a, b) computes the difference between a and b using Manhattan distance as defined below:

$$diff(\mathbf{x}_i^{(A)} - \mathbf{x}_j^{(M)}) = \sum_{k=1}^{n} |x_{i,k}^{(A)} - x_{j,k}^{(M)}| \tag{12}$$

This process is repeated for all instances, and the final feature scores are determined by averaging the accumulated weight updates across the dataset.

2) Phase II - MFO Phase

The Moth Flame Optimization phase refines the feature selection by applying MFO to the ReliefF-selected features.

Moth Flame Optimization (MFO) [41] is a nature-inspired optimization algorithm which mimics the moths flying around flames modelled by transverse orientation mechanism. Moths represent candidate feature subsets, while flames signify the fitness of corresponding subsets. Moths move towards flames using a logarithmic spiral function defined as :

$$X_i^{k+1} = \left\{ \begin{array}{l} \delta_i \cdot e^{bt} \cdot \cos(2\pi t) + Fm_i(k), i \leq n \\ \delta_i \cdot e^{bt} \cdot \cos(2\pi t) + Fm_n(k), i \geq n \end{array} \right\} \tag{13}$$

where, $\delta_i = \left| \vec{X}_i^k - \vec{F}_{m_i} \right|$ denotes how far a moth at position $\vec{X}_i$ is from its corresponding flame ($\vec{F}_{m_i}$) in iteration k, n is the number of flames, t denotes the phase of spiral motion, while b controls the scale of the spiral trajectory, influencing the movement of moths towards flames.

The value of t and n decreases linearly over the iterations, thereby balancing the exploration and exploitation. The mathematical equations for t and n are presented below:

$$r = Cur_{iter} \left(\frac{-1}{Max_{iter}} \right) - 1 \tag{14}$$

$$t = (r - 1) \times k + 1 \tag{15}$$

$$n = round \left(n_{Last\,iter} - Cur_{iter} \frac{(n_{Last\,iter} - 1)}{Max_{iter}} \right) \tag{16}$$

Where, Max_{iter} is the maximum number of iterations, Cur_{iter} is the current iteration, k denotes a random number ranged [0,1], r is the convergence parameter, and its value decreases linearly between [-2,-1].

The algorithm identifies and retains the most promising flames while disregarding the rest, iterating to optimize the feature selection process. These selected features are then utilized in the proposed classification models to ensure robust classification performance.

C. Classification

1) Preactivated Inverted Resnet (PIR)

The proposed classification method demonstrated in fig 4 introduces novel advancements including pre-activated weights in an Inverted-ResNet block with lightweight spatial attention utilizing C-pooling which are then stacked with global average pooling in the Final ResNet node, collectively aiming to improve efficiency and accuracy in image classification. The architecture initiates by integrating palmprint and iris modalities, each processed concurrently within separate branches

Algorithm 2 MFOR: Hybrid Feature Selection Algorithm

Initialize Dataset D, Number of features n, Number of nearest neighbors k, Maximum iterations Max_{iter}
Output: Selected feature subset S
ReliefF Phase:
for $i = 1$ to n **do**
    Compute ReliefF score $W[i]$ for feature i using Eq.(11)
end for
$S_{\text{ReliefF}} \leftarrow$ Features with top k ReliefF scores
MFO Phase:
Initialize n_{flames} moths with random feature subsets
for iter $= 1$ to Max_{iter} **do**
    Update t and n using Eqs. (14), (15), and (16)
    for each moth i **do**
        Compute δ_i
        if $i \leq n$ **then**
            Update X_i using Eq. (13) with $Fm_i = S_{\text{ReliefF}}[i]$
        else
            Update X_i using Eq. (13) with $Fm_n = S_{\text{ReliefF}}[i-n]$
        end if
    end for
    Update n_{flames} by selecting the top moths
end for
$S \leftarrow$ Features corresponding to top moths
return S

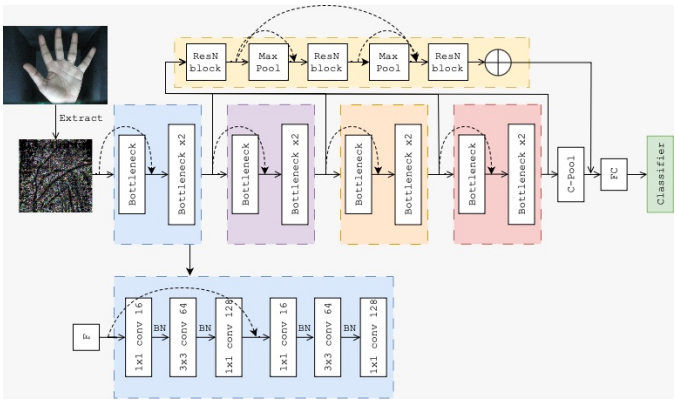


Fig. 4: PIR Architecture

of novel Preactivated Inverted Resnet blocks. The Inverted-ResNet block [42] with pre-activated weights modifies the traditional ResNet architecture by inverting the order of operations within a residual block. Inverted-ResNets with bottleneck connections allow for a highly reduced parameter count. Subsequently, Batch Normalization was used to optimize the high-dimensional network. Pre-activation involves applying activation functions before the skip connection, enhancing stability and convergence, with the leaky ReLU activation function used to prevent the vanishing gradient problem. The leaky Relu function can be defined as :

$$\text{LeakyReLU}(x) = \begin{cases} x, & \text{if } x > 0, \\ \alpha x, & \text{otherwise}, \end{cases} \tag{17}$$

Where α represents a small positive constant (usually around 0.01), which determines the slope of the function for negative input values x.

The output features from the initial ResNet units are split into two branches: the trunk branch and the soft mask branch. The trunk branch ($\mathcal{T}$) processes the features directly, while the soft mask branch generates an attention mask using bilinear interpolation ($\mathcal{F}$) with Resnet block outputs ($\mathcal{U}$).

$$o_{interp} = \mathcal{F}(o_{softmask}) \tag{18}$$

$$o'_{softmask} = \mathcal{U}(o_{interp}) \tag{19}$$

$$o = (1 + o'_{softmask}) \times \mathcal{T}(x) \tag{20}$$

The $o_{softmask}$ is created by downsampling the input tensor through max pooling, passing it through another residual block, and then upsampling back to the original dimensions. The attention mask is applied to the trunk branch output (20), modulating the features to highlight important regions. Finally, the modulated features pass through a final block, ensuring stable and efficient training by leveraging BatchNorm and leaky ReLU activations, and maintaining spatial information through global average pooling.

To facilitate the integration of information from the parallel branches, feature fusion layers are introduced. C-pooling layer is used within each block to combine the feature representations learned from each branch as described below.

$$C_{pooled} = \text{AvgPool}(I \times \text{Att}) \tag{21}$$

Where Att is the attention layer and I is the input. Global average pooling is used to aggregate spatial information across feature maps created by Preactivated Inverted Resnet blocks.

For measuring the model performance of the resnet, the cross-entropy loss function is employed. The cross-entropy function utilizes the softmax activations to generate an output within a probability distribution (p_i). The mathematical representation of cross-entropy loss is:

$$H(p, y) = -\sum_i y_i \log \left(\frac{e^{a_i}}{\sum_{k=1}^{N} e^{a_k}} \right), \tag{22}$$

where $i \in [1, N]$ and $e^{a_i} / \sum_{k=1}^{N} e^{a_k}$ represents the output class probabilities (p_i).

2) Johnson Flower Pollination Algorithm - Rainfall Optimization Algorithm - DenseNet (JFPA-ROA-DenseNet)

Parameter tuning plays a crucial role in the performance of deep learning models. Rosa et al.(2018)[29] highlighted the advantages of using metaheuristics for optimizing these parameters. In the parallel classification model, we propose a novel hybrid metaheuristic approach for optimizing DenseNet. The learning rate and dropout parameter of the DenseNet classifier are optimized through the hybrid of Novel JFPA-ROA. The dropout parameter controls regularization to mitigate overfitting, while the learning rate determines the step size for updating network weights during training.

Johnson Flower Pollination Algorithm (JFPA)

The Flower Pollination Algorithm (FPA) [43] is inspired by natural pollination in flowering plants, where global pollination resembles biotic cross-pollination and promote exploration using the Lévy distribution. A designated flower(pollinator) undergoes a Lévy flight. The updated position of the destination flower is represented as :

$$x_{i+1}^t = x_i^t + \gamma L(g^* - x_i^t) \tag{23}$$

$$L(x) = \begin{cases} 0, & \text{for } x < \mu \\ \sqrt{\frac{\pi}{c\alpha}} \left(\frac{1}{\alpha} \exp\left(-\frac{(\mu - x)^2}{c\alpha} \right) + \frac{1}{1-\alpha} \Gamma(1-\alpha, \frac{x-\mu}{c}) \right), & \text{for } x \geq \mu \end{cases} \tag{24}$$

Where, x_{i+1}^t is the updated solution of ith component at time t, $L(x)$ is the random step size chosen from the probability density function (PDF) of the Lévy distribution, g^* is the global best solution found so far and γ is the scaling parameter which indicates whether the update occurs between neighboring flowers ($\gamma = 1$) or not ($\gamma = 0$). $\Gamma(1-\alpha, \frac{x-\mu}{c})$ is the incomplete gamma function, $\mu(-\infty < \mu < \infty)$ is the location parameter which determines the center of the distribution, $c(c > 0)$ is the scale parameter controlling the spread of the distribution. and the stability parameter α ($0 < \alpha \leq 2$) defines the shape of the tails.

Use of levy distribution would not be able precisely capture the distinct peaks of multimodal biometric data and Its heavy-tailed nature would lead to premature convergence. To overcome these limitations we propose to modify the exploration in FPA by using Johnson distribution as it have 4 tailored parameters leading to broader range of distribution shapes making it more adept at representing the multimodal nature of biometric data.

Definition 3. ***Johnson based Exploration***

The Johnson distribution, a family of probability distributions derived through bijective transformation from the normal distribution, represents a versatile statistical toolset characterized by its probability density function. We incorporate Johnson distribution to update position while exploring the search space in FPA. The modified update position is defined as :

$$x_{i+1}^t = x_i^t + \gamma J(g^* - x_i^t) \tag{25}$$

$$J(x) = \frac{1}{x\sigma\sqrt{2\pi}} \frac{\Gamma \sqrt{\delta^2 + \lambda^2 \left(\frac{x-\xi}{\sigma} \right)}}{\Gamma\left(\sqrt{\delta^2 + \lambda^2} \right)} \tag{26}$$

Where, x_{i+1}^t is the updated solution of ith component at time t,$J(x)$ is the random step size chosen from the PDF of the Johnson distribution σ is the scale parameter, δ and λ are shape parameters and $\Gamma(\cdot)$ is the gamma function.

Local pollination, akin to auto-pollination, occurs probabilistically based on a variable P_c within [0, 1], representing the probability of change. The equation for updating x locally is as follows:

$$x_{i+1}^t = x_i^t + \epsilon \cdot \gamma \cdot (x_j^t - x_k^t) \tag{27}$$

Where, x_t^j and x_t^k represent two randomly selected solutions from the population at iteration t, ϵ is a constant that influences the step size during local pollination.

3) Rainfall Optimization Algorithm (ROA)

The Rainfall Optimization Algorithm (ROA) is a nature-inspired metaheuristic optimization algorithm motivated by the behavior of raindrops flowing down a terrain.

The algorithm initiates with a set of initial raindrops (P), representing a cloud of water droplets, each denoting a potential solution. Initially, all raindrops are considered "active," meaning they participate in the search process.

The algorithm generates neighbor raindrops(N) around the raindrop's current position. Fitness of all active raindrop is calculated as the accuracy of DenseNet classifier. The algorithm updates a raindrop's position to a significantly better neighboring raindrop (Dominant Neighbor), mirroring rain seeking lower elevations. The update equation is as follows:

$$x_i^{(t+1)} = Max_{j \in \{1,2,\ldots,n\}}(F(x_j^{(t)})) \tag{28}$$

Where $x_i^{(t)}$ is the position of raindrop i at iteration t and $j \in \{1, 2, ..., n\}$ represent the neighbor raindrops.

If no clear advantage is found, the "explosion process" generates random new neighbor raindrops, akin to rain dispersing upon hitting the ground, fostering exploration in diverse directions.

$$x_i^{(t+1)} = \begin{cases} x_i^{(t)} + E_p & \text{if } F(x_i^{(t)} + E_p) > F(x_i^{(t)}) \\ x_i^{(t)} & \text{otherwise} \end{cases} \tag{29}$$

Where, E_p is the explosion point calculated using johnson distribution.

After a predefined number (Ne) of unsuccessful explosion attempts, the raindrop is marked as "inactive." This prevents the algorithm from getting stuck in areas with low potential for improvement, similar to how stagnant water "evaporates". After processing all active raindrops, raindrops are ranked by fitness values. Low-ranking ones may be removed for diversity or high-ranking ones given more attempts. The loop continues until all raindrops are inactive or maximum iterations are reached, then raindrop with best fitness is selected.

4) JFP-ROA-DenseNet

JFPA's focus on global exploration may limit its effectiveness in exploiting local optima, potentially hindering its performance in complex multimodal image recognition tasks. Rainfall Optimization Algorithm excels in local search and can help FPA fine-tune its solutions by efficiently exploiting nearby regions. ROA may excel in local search but may not provide adequate global exploration. To address these limitations, our research proposes a hybrid approach that integrates ROA into JFPA for local pollination, targeting the optimization of DenseNet's dropout and learning rate and dropout parameters, thereby enabling enhancements in transfer learning-based DenseNet (TL-JFPA-ROA-DenseNet). This hybridization aims to strike a balance between global and local search, enhancing the algorithm's capabilities for improved performance.

Algorithm 3 illustrates the application of the TL-JFPA-ROA-DenseNet algorithm for optimizing the learning rate and dropout parameter within the DenseNet classifier.

Algorithm 3 TL-JFPA-ROA-DenseNet

Require: P: Population, Max_iter: Max iterations, P_c: Probabilty of change

for Each flower i in P **do**
  Evaluate the fitness of flower $F(i)$
  if $F(i) \geq F(f_{best})$ **then**
    $f_{best} \leftarrow i$
  end if
end for
for $t = 0$ in Max_iter **do**
  Generate random steps by *Johnson Distribution* eq(26)
  for Each flower i in P **do**
    Potential Flower $f_p \leftarrow i$
    Apply random steps to f_p *Global Pollination*
    if $P_{rand} > P_c$ **then**
      Use *ROA* for *Local Pollination*
    end if
    Calculate Fitness $F(i)$
    if $F(f_p) > F(i)$ **then**
      $i \leftarrow f_p$
      $F(i) \leftarrow F(f_p)$
      if $F(f_p) > F(f_{best})$ **then**
        $f_{best} \leftarrow f_p$
      end if
    end if
  end for
end for
return f_{best}

D. Fusion

For fusion of results from both the classification approaches Score-level fusion is adopted. To combine scores from M different modalities $(S_1, S_2, ..., S_M)$ into a single score vector S, a fusion function F is employed. The mathematical representation of the product rule of fusion used is as follows:

$$\begin{aligned} S &= F(S_{PIR}, S_{JFPA-ROA-DenseNet}) \\ &= \alpha * S_{PIR} + (1-\alpha) * S_{JFPA-ROA-DenseNet} \end{aligned} \tag{30}$$

The final score vector S is used to deduce results.

V. Experiment

In this section, we first outline the experimental setup essential for implementing our proposed methodologies. Following this, we provide a detailed explanation of implementing each proposed approach. This includes step-by-step descriptions from image preprocessing and feature selection to the implementation of our classifiers and their fusion. This approach ensures a comprehensive understanding of how each element contributes to our methodology.

A. Experimental Setup

The proposed methodolgy was put into implementation using Python 3 within the Google Colab environment that offers a convenient platform for running Python code with access to various hardware resources such as CPU, TPU (Tensor Processing Unit), or T4-GPU, depending on availability and requirements for speeding up computation of deep learning models and metaheuristic optimization techniques. Additionally, the system was equipped with 12.7 GB of RAM , providing sufficient memory for processing and storing data during the execution of the proposed methodology thus avoiding performance bottlenecks due to memory limitations. Overall, leveraging Python 3 on Google Colab, along with CPU, TPU, or T4-GPU runtimes, and supported by 12.7 GB of system RAM, ensured an optimal environment for implementing and executing the proposed methodology effectively and efficiently.

B. Implementation

In this subsection, we will outline the implementation of our proposed methodology. The process is detailed in the following points

- **Image Preprocessing**:
  The biometric image is initially enhanced using the CLAHE algorithm, with the determination of its Clip Limit parameter by using the novel MFALF algorithm.The image is initially resized to dimensions of 224 x 224 pixels utilizing OpenCV, converted to black and white, and subsequently forwarded to the CLAHE algorithm. In order to predict the Clip Limit, the search space for the CLAHE algorithm's Clip Limit parameter ranges from 0 to 60. The fireflies' population size is fixed at 10, with a total of 100 iterations set to find the optimal Clip Limit for image enhancement. For the CASIA Palmprint dataset, the optimal Clip Limit obtained was 18.67, while for MMU Iris image it is 20. In order To evaluate the performance of the proposed MFALF algorithm, we also applied well-known metaheuristic algorithms, including the Firefly Algorithm, Particle Swarm Optimization (PSO), Whale Optimization Algorithm (WOA), and Cat Swarm Optimization Algorithm (CSOA) for image enhancement in biometric recognition. The fitness function combines PSNR and SSIM, aiming to minimize it to extract useful features during image enhancement. As shown in Table I , the proposed MFALF algorithm achieves the lowest value (3.076) on CASIA palmprint image, outperforming the other metaheuristic algorithms in determining the most optimal Clip Limit for the CLAHE algorithm.

TABLE I: MFALF Performance Comparison

Algorithm	Clip Limit	Fitness Function
PSO	9.85	7.687
CSOA	2	7.617
WOA	60	5.264
Firefly	35.5	5.107
MFALF	**20**	**3.076**

- **Feature Selection**:
  After enhancing the contrast of the images, a novel hybrid methodology integrating ReliefF and Moth Flame Optimization (MFO) is introduced for the selection of informative features. This approach is structured into two distinct phases.

In the first phase, the initial set of 150,528 features undergoes processing via the ReliefF algorithm, which is a filter method that assesses the relevance of features independently of any specific learning algorithm. For this procedure, the number of neighbors is fixed at 100, which determines the number of neighbors considered when assigning feature importance scores; a higher number of neighbors leads to more accurate scores, albeit at the cost of increased computation time. The 'n_features_to_keep' parameter, which specifies the number of features to be retained post-ReliefF processing, is maintained as a variable to allow for adjustment based on specific requirements.

Subsequently, in the second phase, the features selected by the ReliefF algorithm are further refined using Moth Flame Optimization (MFO). MFO functions as a wrapper method, evaluating subsets of features based on the performance of a particular learning algorithm. In this phase, the maximum number of iterations is set to 10, and the population size is established at 20, representing the number of moths (candidate solutions) in the population. The parameter determining the final number of required features is maintained as a variable. An SVM classifier is employed to compute the fitness and accuracy of the feature subsets during this phase.The SVM classifier is preferred for its robustness in high-dimensional spaces and effectiveness with small to medium-sized datasets, making it ideal for the refined feature subsets in this methodology.

- **JFPROA Densenet**:
  We receive biometric image dataset with selected features through the feature selection process, which serves as the foundation for training and testing our novel model. Our approach focuses on optimizing the hyperparameters of a pre-trained DenseNet-121 model for biometric image classification. This optimization is achieved using a combined Johnson Flower Pollination Algorithm and Rainfall Optimization Algorithm (JFPA-ROA). We begin by refining the model's classifier with enhancements such as batch normalization, ReLU activations, and adaptive dropout layers. The process starts with a random initialization of a diverse set of hyperparameter configurations, including parameters like learning rate, & dropout rate. Key parameters such as population size (set to 2), maximum iterations (set to 2), and beta (set to 0.5) are predefined to balance exploration and exploitation within the algorithm. Specific ROA parameters (gamma = 2.0, delta = 3.0, epsilon = 0.01) shape the Johnson distribution used for random updates during exploration. Each configuration undergoes training to assess its performance based on validation accuracy. Through iterative refinement employing global pollination and explosion processes, the algorithm dynamically adjusts hyperparameters to optimize model performance. Learning rate & dropout parameter determined by the proposed JFPROA algorithm is 0.001 & 0.205 . Hence, in the pretrained Densenet121 CNN model, we set the Learning rate parameter to 0.001 and dropout rate parameter to 0.0205 and train out biometric images over the optimized model. This methodology not only highlights the efficacy of bio-inspired optimization techniques in deep learning but also facilitates the discovery of optimized model settings tailored for intricate image classification tasks
- **PIR**: We have designed a novel PIR architecture to function as a secondary classifier alongside the JFPROA DenseNet in a multimodal biometric recognition system. In this multimodal system, both classifiers operate simultaneously. For instance, if palmprint images are trained using the transfer learning-based JFPROA DenseNet, then iris images are trained and tested using the PIR architecture, and vice versa. The PIR Architecture uses a pre-activated inverted ResNet with a spatial attention mechanism. We optimize our approach over a residual convolutional neural network, leveraging the efficiency of existing inverted resnets (for eg. MobileNetv2), using a combination of lightweight stacked layers for spatial attention with C-pooling to improve the generalizability of models trained on small datasets. This setup enhances the network's focus on critical regions within an image and improves feature capture. By incorporating BatchNorm and pre-activation with leaky ReLU, we stabilize and accelerate the training process, leading to better gradient flow and reduced vanishing gradient problems. Our final architecture employs global average pooling to minimize parameters while maintaining spatial information, resulting in a robust, efficient, and scalable solution for biometric classification.

  PIR consists of multiple dense blocks where each layer is connected to every other layer in a feed-forward fashion. The preprocessed biometric images are then passed through the initial convolution and pooling layers of the network, which help in capturing basic features like edges and vein textures, and a lightweight stack of layers for spatial attention using C-pooling, enabling spatial feature extraction. The output features from the intial attention mechanism are subsequently fed into the inverted ResNet blocks with pre-activated weights. Each block contains several convolutional layers, where each layer benefits from pre-activation with leaky ReLU and BatchNorm, ensuring stable and efficient training.

  By using pre-activation, the activation functions are applied before the weights, which improves gradient flow and reduces the vanishing gradient problem. This allows the network to effectively train deeper architectures without degradation. Batch normalization further stabilizes the training process by normalizing the inputs to each layer, reducing internal covariate shift and ensuring consistent input distributions that prevent the model from overfitting to unique and complex patterns in the images. This combination of pre-activation, leaky ReLU, and BatchNorm leads to a more robust and efficient training process, which we note empirically in section VI.

  Each resnet block incorporates additional residual information provided by the attention block before passing on the normalized weights. We also find that incorporating an additional BatchNorm layer after the residual

operation further improves performance over the select datasets. The output from the final ResNet block is then passed through a global average pooling (C-pooling) layer, reducing the dimensions of the feature maps while maintaining spatial information. To further enhance the classifier, we find that including adaptive dropout layers (dropout rate $p = 0.2$) helps in regularizing the model and reducing the variability and noise in the datasets. Finally, the reduced feature maps are fed into a fully connected layer for classification into the specified number of classes. The PIR model was initialized with the desired number of output classes and trained using Adam optimization. The performance of the trained model was evaluated on separate validation and test sets to assess classification accuracy, convergence behavior, and computational efficiency. These enhancements and modifications enabled the development of a robust, efficient, and scalable framework tailored specifically for biometric recognition tasks.

- **Fusion**:
  To evaluate the multimodal biometrics dataset (palmprint & iris) results, we used a basic score level fusion approach to determine final scores. Using the final scores achieved, we assign the class with the highest score to the particular test image. We experimented to find the best alpha value in Eq 30 for every multimodal dataset M-BIRJ, M-IIT, M-CAS.

In the upcoming section VI, we will delve into an analysis of how effectively our proposed classifiers perform. This examination will cover various performance metrics and provide a comprehensive evaluation of their strengths and weaknesses.

VI. CLASSIFICATION RESULTS

In this section, we evaluate the performance of the proposed methodologies using key metrics, including Accuracy (ACC), Equal Error Rate (EER), False Acceptance Rate (FAR), False Rejection Rate (FRR), Total Training Time, Macro F1-Score, Detection Error Tradeoff (DET) Curve, and Receiver Operating Characteristic (ROC) Curve. These metrics collectively provide a robust framework for assessing the effectiveness of the novel approach for classification. This includes evaluating the impact of image preprocessing and feature selection, comparing the approach with other CNN architectures, comparing unimodal and multimodal biometric datasets,and assessing the overall versatility of the proposed approach alongside a comparison of the proposed classifiers. This comprehensive evaluation enhances understanding and guides future advancements in the field.

It is important to note that for evaluating the results of the unimodal datasets, IIT palmprint images are trained using Novel Preactivated Inverted Resnet architecture (PIR) , while MMU iris images, CASIA & BIRJAND palmprint images are trained using proposed Transfered Learning based JFP-ROA Densenet (JFPROA-Densenet) CNN Model.

A. Impact of Image Preprocessing

The impact of preprocessing on the performance of our proposed model is highlighted in Table II. This illustrates the impact of preprocessing on the performance metrics of the proposed model with the number of features selected in the range of 65k-75k. It is evident from Table II that preprocessed biometric images of all unimodal dataset (MMU, CASIA, BIRJAND, IIT) outperform unprocessed images across various metrics including Accuracy, FAR, FRR, and Macro F1 score, all achieved within a shorter total training time. This demonstrates the effectiveness of novel preprocessing approach in reducing the computational requirements & enhancing the performance of biometric recognition systems.

B. Impact of Feature Selection

Now let us understand the impact of feature selection on performance metrics of the proposed classifiers for various preprocessed unimodal biometric datasets. Table III showcases the performance of the proposed model for MMU , CASIA, BIRJ unimodal datasets, both without applying the MFOR algorithm and with the application of the MFOR algorithm for selecting the number of features within the ranges of 65k-75k and 35k-50k. It is evident that when the number of selected features falls within the range of 65k-75k, the model achieves maximum accuracy, Macro F1 score, and lower FAR and FRR values, all while reducing the total training time. This highlights the efficiency of our proposed MFOR algorithm for feature selection, which effectively selects relevant features, reduces computational complexity and risk of overfitting, while enhancing accuracy and other performance metrics of the proposed approach.

C. Performance of Preprocessed Unimodal Biometric Recognition using Different Models

As previously mentioned, for the results of unimodal datasets (as shown in Table II and Table III), such as the IIT Palmprint dataset, we utilized the proposed Preactivated Inverted Resnet (PIR) CNN architecture for training, while MMU Iris images, CASIA & BIRJAND Palmprint images were trained over the novel Transfer Learning based JFP-ROA Densenet Model.

To assess and compare the performance of the proposed classifiers, we accomplish training of MMU iris and IIT palmprint datasets over popular pre-trained CNN architectures like InceptionV3, Resnet50 as presented in Table IV and Table V. Analysis of Table IV reveals that our proposed models (JFP-ROA Densenet and PIR) for classification exhibit superior performance in terms of accuracy, train loss, and val loss achieved within a shorter duration of total training time for the MMU Iris dataset, demonstrating its practical deployment potential.

In the analysis of the IIT Palmprint dataset presented in Table V, the trends are similar to those observed in Table IV. However, when comparing the performance of our proposed JFP-ROA Densenet model and Resnet50, we notice that the train loss for the proposed Densenet (0.0021) is higher than that of Resnet50(0.0007). Nevertheless, the validation loss(Val Loss) of JFP-ROA Densenet (0.0082) is lower as compared to Resnet50 (0.0095), indicating that our proposed Densenet model avoids overfitting and achieves better accuracy within

TABLE II: Dataset Preprocessing Impact

Dataset	Preprocessing	Clip Limit	ACC	FAR	FRR	Macro F1 Score	Total Training Time
MMU	**YES**	**20**	**0.9988**	**0**	**0.0011 ± 0.0737**	**0.9997**	**21 min 33 sec**
	NO	-	0.8684	0.1306	0.0018	0.8177	48 min
CASIA	**YES**	**18.67**	**0.9875**	**0**	**0.0075 ± 0.04967**	**0.9734**	**32 min 8 sec**
	NO	-	0.8453	0.1896	0.0176	0.7983	46 min n
BIRJAND	**YES**	**19.34**	**0.9938**	**0.01780 ± 0.00105**	**0.5555 ± 0.0328**	**0.99477**	**3 hr 51 min 5 sec**
	NO	-	0.886666	0.3148747 ± 0.005888	12.181467 ± 0.221159	0.8595663	2 hr 20 min
IIT	**YES**	**20**	**0.96956**	**3.8732 ± 0.178747**	**0.01544 ± 0.000962**	**0.9436**	**69 min 15 sec**
	NO	-	0.8632	0.1487 ± 0.00041	0.1523 ± 0.0289	0.8434	75 min 20 sec

TABLE III: Feature selection impact on performance

Dataset	Range of Features	Features Selected	ACC	FAR	FRR	Macro F1 Score	Total Training Time	Training Time per Epoch
MMU	-	1,50,528	0.98684	0.00039	0.00952	0.9902	28 min 3 sec	3.1 min
	65k-75k	**74,413**	**0.9988**	**0**	**0.0011 ± 0.0737**	**0.9997**	**21 min 33 sec**	**2.08 min**
	35k-50k	46,181	0.98684	0.0004	0.0095	0.9902	29 min 41 sec	2.9 min
CASIA	-	1,50,528	0.9801	0.02436 ± 0.00312	0.785 ± 0.01237	0.9701	34 min 19 sec	5.23 min
	65k-75k	**70,272**	**0.9875**	**0**	**0.0075 ± 0.0497**	**0.9734**	**32 min 8 sec**	**4.57 min**
	35k-50k	45,219	0.9611	0.01436 ± 0.00312	0.985 ± 0.01237	0.9701	35 min 41 sec	5.15 min
BIRJAND	-	1,50,528	0.9938	0.01736 ± 0.00102	0.5555 ± 0.03286	0.99135	4 hr 5 min 30 sec	15 min 30 sec
	65k-75k	**67,795**	**0.9938**	**0.01780 ± 0.00105**	**0.5555 ± 0.0328**	**0.99477**	**3 hr 51 min 5 sec**	**14 min 26 sec**
	35k-50k	44,956	0.9382	0.15890 ± 0.00406	4.4598 ± 0.1100	0.9250	3 hr 45 min	15 min

a much shorter duration (39 min) of total training time compared to Resnet50 (70 min).

Hence, We can conclude that our proposed models presents superior result as compared to other pretrained CNN architectures. And so the scores from the novel JFP-ROA Densenet architecture are fused with the scores of the proposed Preactivated Inverted Resnet (PIR) for evaluating the multimodal biometric results, as depicted in Table VIII.

We have evaluated the performance of our proposed classifier using different optimizers like Adam & RMSProp. Selecting the appropriate optimizer for a CNN model is imperative as it significantly impacts training speed, generalization, robustness, efficiency, hyperparameter sensitivity, and task-specific performance.Adam adapts learning rates per parameter, ideal for dynamic and noisy datasets, while RMSprop efficiently handles non-stationary objectives and converges faster on convex problems. Table VI presents the performance of our proposed model across various unimodal datasets with different optimizers. From the analysis of Table VI , it is evident that superior results are achieved when the Adam optimizer is used compared to the RMSPROP optimizer. Consequently, we have opted to use the Adam optimizer for our proposed classification models for training noisy biometric images.

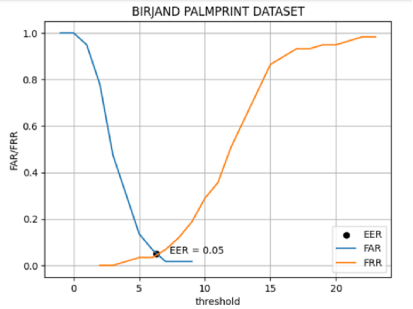
We will now explore the rationale behind adopting a transfer learning approach for our model, as assessed in Table VII . This table presents a comparative analysis of test results with and without transfer learning across various unimodal datasets trained using our proposed methodology Transfer Learning based JFPROA Densenet. It underscores the substantial performance improvements and enhanced stability achieved through transfer learning in biometric recognition. By leveraging knowledge transfer from related tasks, transfer learning effectively enhances both the efficacy and computational efficiency of deep learning models, thereby revolutionizing their training paradigm.

D. Performance Comparision of Unimodal and Multimodal dataset

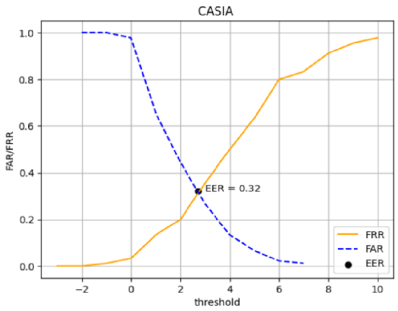
Now, let's compare the performance of our proposed classifiers on preprocessed unimodal datasets and preprocessed multimodal datasets, where the number of selected features ranges from 65k to 75k. Table VIII presents the comparison of performance of the M-CAS, M-BIRJ, M-IIT, and M-CAS2 multimodal datasets with the Unimodal dataset. In the M-CAS & M-IIT dataset, MMU Iris images are trained using the novel Transfer Learning based JFP-ROA-Densenet architecture, while CASIA & IIT palmprint images are trained using the proposed PIR CNN architecture. In contrast, for the M-BIRJ & M-CAS2 multimodal dataset, MMU Iris images are trained using the novel PIR, while BIRJ & CASIA palmprint images are trained using the proposed Transfer Learning based JFP-ROA Densenet model.

Analysis of Table VIII indicates that the accuracy of the M-CAS multimodal dataset is 1, which surpasses the accuracy achieved by the corresponding unimodal datasets: CASIA Palmprint (0.9334) and MMU Iris (0.9988). Similar trends are observed for other multimodal datasets such as M-BIRJ, M-IIT, and M-CAS2. The performance of multimodal datasets, as measured by other metrics like FAR, FRR, and F1-score, also exhibits superiority compared to the performance of unimodal datasets. This highlights the effectiveness of leveraging multiple modalities (palmprint and iris) in biometric recognition, leading to enhanced accuracy and performance across various metrics.

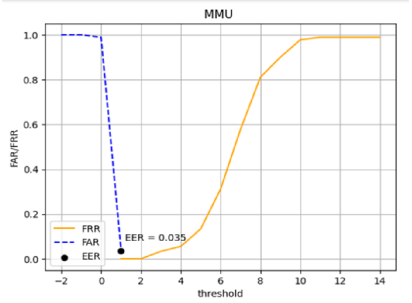
To assess the performance of the multimodal M-BIRJ dataset, we determine the final class prediction score by averaging the scores predicted by the Transfer Learning based JFP-ROA Densenet (for palmprint) and the PIR model (for iris). Subsequently, the test image is assigned to the class with the highest score. Through experimentation, we established that setting alpha to 0.5 yields optimal accuracy (1). Similarly, in the case of the M-IIT dataset, the peak accuracy(1) is attained when alpha equals 0.5. Conversely, for M-CAS multimodal



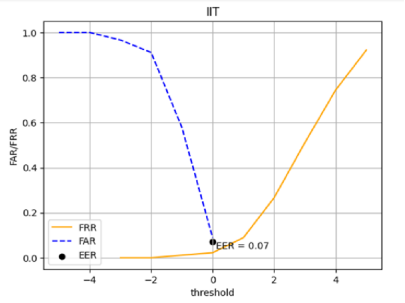
(a) Birjand Dataset EER Graph



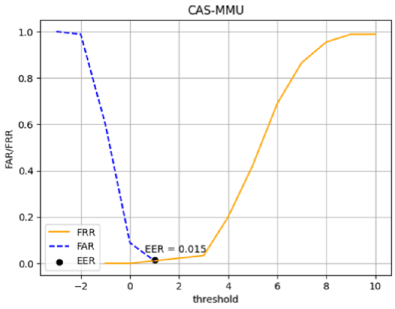
(b) Casia Dataset EER Graph



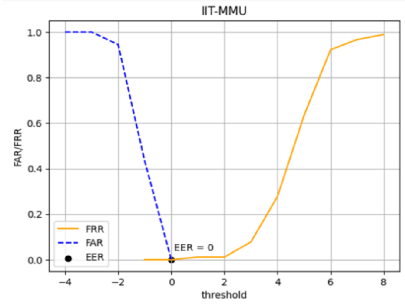
(c) MMU Dataset EER Graph



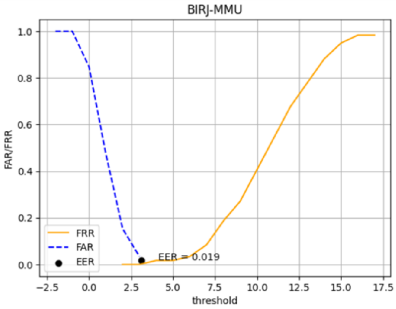
(d) IIT Dataset EER Graph



(e) M-CAS Dataset EER Graph



(f) M-IIT Dataset EER Graph



(g) M-BIRJ Dataset EER Graph

Fig. 5: Equal Error Rate Graphs

TABLE IV: Performance Metrics of Different Models On MMU with 74,413 selected features

Model	ACC	Train Loss	Val Loss	Total Training Time	Training Time per Epoch	Input Size
InceptionV3	0.7681	0.0001	0.0154	33 min	5.5 min	2,68,203
PIR	**0.9605**	**0.0001**	**0.0034**	**15 min**	**55 sec**	**1,50,528**
JFP-ROA Densenet	**0.9988**	**0.0001**	**0.0006**	**21 min 33 sec**	**2.08 min**	**1,50,528**
ResNet50	0.9464	0.0012	0.0078	26.4 min	2.4 min	1,50,528

dataset, accuracy reaches 1 when alpha is set to 0.6.

FAR (False Acceptance Rate) refers to mistakenly accepting an imposter as genuine, while FRR (False Rejection Rate) occurs when a genuine person is wrongly rejected. EER (Equal Error Rate) marks the point where the difference between FAR & FRR is minimum . At this threshold, decisions are made: scores below it are classified as imposters, and scores above it as genuine.Table IX indicates that the EER for multimodal datasets is significantly lower compared to unimodal datasets, demonstrating superior performance.

The FAR/FRR vs Threshold graph (Fig 5) helps in evaluating the EER and decision threshold. The performance of both unimodal and multimodal datasets were evaluated using the Detection Error Tradeoff(DET) curve, which is a graphical representation that illustrates the relationship between FAR and FRR. Fig 6 demonstrates that the FAR-FRR tradeoff for the M-BIRJ multimodal dataset is better than that of the unimodal MMU and BIRJAND dataset. Similar Graphs are observed for other multimodal datasets(M-CAS & M-IIT) as shown in Fig 6 . Receiver Operating Curve (ROC) depicts the tradeoff between True Positive Rate (TPR) and False Positive Rate(FPR) . Micro-Average ROC curves for CASIA & M-CAS datasets are shown in Fig 7. While ROC curves are commonly used, DET curves offer finer control over error rates, crucial for biometric recognition assessment. Hence, DET curves are often favored for comprehensive biometric performance analysis.

TABLE V: Performance Metrics of Different Models On IIT dataset with 67,795 selected features

Model	ACC	Train Loss	Val Loss	Total Training Time	Training Time per Epoch	Input Size
InceptionV3	0.80434	0.0060	0.5914	100 min 12 sec	8 min 32 sec	(3,299,299)
PIR	**0.96956**	**0.0003**	**0.0044**	**69 min 15 sec**	6 min 56 sec	(3,224,224)
JFP-ROA Densenet	**0.93913**	0.0021	**0.0082**	**39 min 30 sec**	5 min 34 sec	(3,224,224)
ResNet50	0.93043	**0.0007**	0.0095	70 min 12 sec	6 min 14 sec	(3,224,224)

TABLE VI: Impact of Different Optimizers for Unimodal datasets

Dataset	Optimizer	Runtime Type	ACC	FRR	FAR	Macro F1 Score	Total Training Time	Training Time per Epoch
BIRJAND	**Adam**	CPU	**0.9938**	**0.01780 $\pm$ 0.00105**	**0.5555 $\pm$ 0.0328**	**0.9947**	**3 hr 51 min**	**12 min 7 sec**
	RMSProp	CPU	0.5061728	51.95105 $\pm$ 0.408361	1.430467 $\pm$ 0.032392	0.4281392	4 hr 28 min	14 min 30 sec
IIT	**Adam**	CPU	**0.96956**	**3.8732 $\pm$ 0.178747**	**0.01544 $\pm$ 0.000962**	**0.9436**	**69 min 15 sec**	**6 min 56 sec**
	RMSProp	CPU	0.76086	25.5516 $\pm$ 0.405105	0.117156 $\pm$ 0.002888	0.6333	104 min	6 min 56 sec
CASIA	**Adam**	CPU	**0.9875**	**0.0075 $\pm$ 0.04967**	**0**	**0.9734**	**32 min 8 sec**	**4.57 min**
	RMSProp	CPU	0.6375	0.2841 $\pm$ 0.3885	0.0078 $\pm$ 0.0175	0.6484	37 min 5 sec	5.28 min
MMU	**Adam**	T4-GPU	**0.9988**	**0**	**0.0011 $\pm$ 0.0737**	**0.9997**	**21 min 33 sec**	**2.08 min**
	RMSProp	T4-GPU	0.8026	0.18	0.00422	0.7468	36 min 59 sec	2.47 min

TABLE VII: Impact of Transfer Learning for Unimodal datasets

Dataset	Transfer Learning	Runtime Type	ACC	FAR	FRR	Macro F1 Score	Total Training Time
MMU	**YES**	T4-GPU	**0.9988**	**0**	**0.0011 $\pm$ 0.0737**	**0.9997**	**21 min 33 sec**
	NO	T4-GPU	0.986	0.004	0.0018	0.967	48 min
CASIA	**YES**	CPU	**0.9875**	**0**	**0.0075 $\pm$ 0.04967**	**0.9734**	**32 min 8 sec**
	NO	CPU	0.6125	0.0034 $\pm$ 0.0101	0.3522 $\pm$ 0.444	0.594	35 min
BIRJAND	**YES**	TPU	**0.98765**	**0.0349 $\pm$ 0.0014**	**1.25 $\pm$ 0.051874**	**0.9843**	**111 min**
	NO	TPU	0.8827	0.2831 $\pm$ 0.0043	9.4290 $\pm$ 0.1517	0.8769	155 min

TABLE VIII: Multimodal Vs Unimodal Performance Comparison

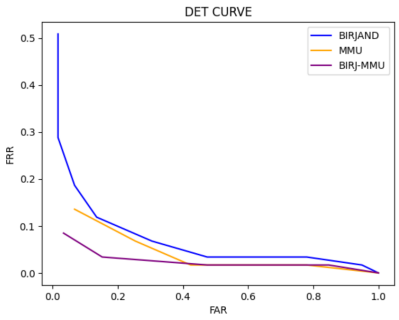
Dataset		ACC	FRR	FAR	F1-Score
M-CAS	**M-CAS**	**1**	**0**	**0**	**1**
	MMU	0.9988	0	0.0011 ± 0.0737	0.9997
	CASIA	0.9334	0.023±0.0123	0.075± 0.049	0.9439
M-IIT	**M-IIT**	**1**	**0**	**0**	**1**
	MMU	0.9988	0	0.0011 ± 0.0737	0.9997
	IIT	0.9695	3.873±0.17	0.015±0.00096	0.9436
M-BIRJ	**M-BIRJ**	**1**	**0**	**0**	**1**
	MMU	0.9491	0.0308 ± 0.1114	0.00196 ± 0.0074	0.9594
	BIRJ	0.9938	0.01780 ± 0.00105	0.5555 ± 0.0328	0.99477
M-CAS2	**M-CAS2**	**1**	**0**	**0**	**1**
	MMU	0.9491	0.0308 ± 0.1114	0.00196 ± 0.0074	0.9594
	CASIA	0.9875	0	0.075± 0.049	0.9734

E. Comparison between the Proposed Classifiers & Versatility of our Proposed Approaches

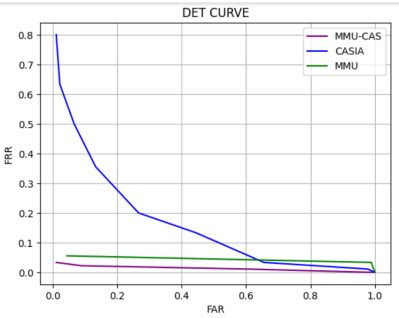
Finally, we conduct a comprehensive comparison between our two novel CNN models, emphasizing their versatility to be trained on either modality. This showcases the adaptability and robustness of our models across different biometric data configurations.

As observed in Table IV & V, our proposed classifying models outperform other CNN classifiers. However, examining Table X, we notice that our proposed Transfer Learning based JFP-ROA Densenet model provides better accuracy for CASIA, MMU, and Birjand datasets compared to the proposed Preactivated Pre-trained Inverted ResNet. Nonetheless, it is essential to note that the Preactivated Pre-trained Inverted ResNet trains much faster. While the former algorithm offers superior results, it requires comparatively more training time. On the other hand, the preactivated inverted ResNet algorithm achieves good accuracy in significantly less training time. Therefore, the score-level fusion of both models helps enhance each other's performance, ultimately leading to the best performance for multimodal biometric recognition.

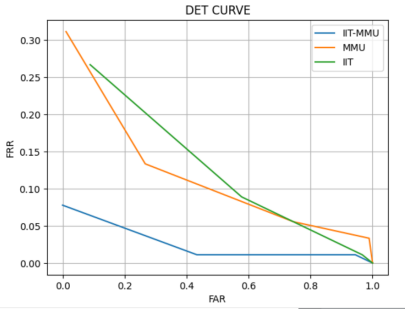
In Table VIII, comparing the results of M-CAS, where MMU iris dataset is trained using the Novel Transfer Learning based JFP-ROA Densenet model, and innovative PIR Model is used for training CASIA palmprint images, with M-CAS2 multimodal biometric dataset, where MMU iris images are trained using Novel PIR and CASIA images are trained using the proposed JFP-ROA Densenet model, both achieve



(a) M-BIRJ DET Curve



(b) M-CAS DET Curve



(c) M-IIT DET Curve

Fig. 6: Detection Error Tradeoff Graphs

TABLE IX: Equal Error Rate (EER) and Threshold values for multiple datasets

Dataset	Model	EER	Decision Threshold	ACC
CASIA	PIR	0.32	2.7	0.9334
MMU	PIR	0.12	4.7	0.9491
MMU	JFPA-ROA-DenseNet	0.035	0.9	1
IIT	PIR	0.07	0	0.9782
BIRJAND	JFPA-ROA-DenseNet	0.05	6.3	0.9941
M-CAS	**PIR + JFPA-ROA-DenseNet**	**0.015**	**1**	**1**
M-BIRJ	**PIR + JFPA-ROA-DenseNet**	**0.019**	**3.1**	**1**
M-IIT	**PIR + JFPA-ROA-DenseNet**	**0**	**0**	**1**

an accuracy of 1. This demonstrates the versatility of our proposed classifiers, showing its effectiveness across different modalities to deliver the best multimodal results.

VII. Conclusion

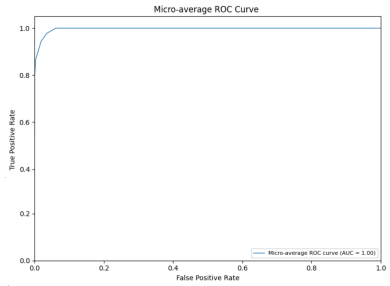
In this paper, we propose an innovative multimodal biometric recognition system leveraging deep learning techniques with iris and palmprint modalities. Our research introduces several novel methods to enhance efficiency and accuracy. Initially, the Modified Firefly Algorithm with Lévy Flights (MFALF) optimizes the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm, enhancing image contrast. Following this, a unique hybrid of ReliefF and Moth Flame Optimization (MFOR) is employed for feature selection, extracting informative features.

For classification, a parallel approach is adopted. Firstly, a novel Preactivated Inverted ResNet (PIR) architecture is introduced. Secondly, metaheuristics are utilized, integrating Johnson Flower Pollination Algorithm with Rainfall Optimization Algorithm for fine tuning of the learning rate and dropout parameters of Transfer Learning based DenseNet architecture (JFPROA DenseNet). Finally, a score-level fusion strategy combines the outputs of the classifiers, ensuring robust and accurate multimodal biometric recognition.

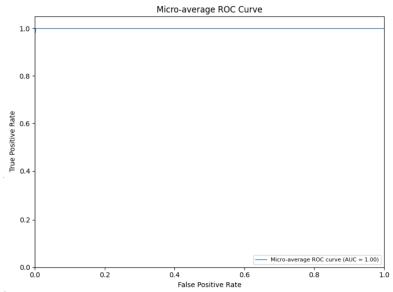
This multimodal approach achieves 100% recognition accuracy on various datasets, surpassing traditional unimodal iris and palmprint identification methods. It demonstrates high efficacy with a lower Equal Error Rate (EER), False Acceptance Rate (FAR), and False Rejection Rate (FRR). The Detection

TABLE X: Performance comparision between JFPROA Densenet & PIR

Dataset	Accuracy		Total Training Time (minutes)	
	JFPA-ROA Densenet	PIR	JFPA-ROA Densenet	PIR
CASIA	**0.9875**	0.9334	32	**19**
MMU	**0.9988**	0.9491	**21**	22
BIRJAND	**0.9938**	0.9875	231	**186**
IIT	0.9574	**0.9695**	73	**69**



(a) CASIA ROC Curve



(b) M-CAS ROC Curve

Fig. 7: ROC Graphs

Error Tradeoff (DET) curve further confirms the system's superiority over unimodal recognition. While highly scalable for a diverse dataset with a large number of individuals, the system's scalability is ultimately limited by computational resources and task complexity. Future improvements can focus on incorporating additional biometric traits like finger vein, palm vein, and facial features. This would further reduce the EER and enhance the overall recognition accuracy, creating an even more robust and secure multimodal biometric system

REFERENCES

[1] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[2] Shaymaa Adnan Abdulrahman and Bilal Alhayani. A comprehensive survey on the biometric systems based on physiological and behavioural characteristics. *Materials Today: Proceedings*, 80:2642–2646, 2023.

[3] Kevin W Bowyer, Karen Hollingsworth, and Patrick J Flynn. Image understanding for iris biometrics: A survey. *Computer vision and image understanding*, 110(2):281–307, 2008.

[4] Adams Kong, David Zhang, and Mohamed Kamel. A survey of palmprint recognition. *pattern recognition*, 42(7):1408–1418, 2009.

[5] Muhtahir O Oloyede and Gerhard P Hancke. Unimodal and multimodal biometric sensing systems: a review. *IEEE access*, 4:7532–7555, 2016.

[6] Shervin Minaee, Amirali Abdolrashidi, Hang Su, Mohammed Bennamoun, and David Zhang. Biometrics recognition using deep learning: A survey. *Artificial Intelligence Review*, 56(8):8647–8695, 2023.

[7] Nour Elhouda Chalabi, Abdelouahab Attia, Abderraouf Bouziane, Mahmoud Hassaballah, and Zahid Akhtar. Recent trends in face recognition using metaheuristic optimization. In *Handbook of nature-inspired optimization algorithms: the state of the art: volume II: solving constrained single objective real-parameter optimization problems*, pages 85–112. Springer, 2022.

[8] Alfred C Weaver. Biometric authentication. *Computer*, 39(2):96–97, 2006.

[9] Ritesh Vyas, Tirupathiraju Kanumuri, Gyanendra Sheoran, and Pawan Dubey. Accurate feature extraction for multimodal biometrics combining iris and palmprint. *Journal of Ambient Intelligence and Humanized Computing*, 13(12):5581–5589, 2022.

[10] Zhenan Sun, Libin Wang, and Tieniu Tan. Ordinal feature selection for iris and palmprint recognition. *IEEE Transactions on Image Processing*, 23(9):3922–3934, 2014.

[11] Nada Alay and Heyam H Al-Baity. Deep learning approach for multimodal biometric recognition system based on fusion of iris, face, and finger vein traits. *Sensors*, 20(19):5523, 2020.

[12] Nema Salem, Hebatullah Malik, and Asmaa Shams. Medical image enhancement based on histogram algorithms. *Procedia Computer Science*, 163:300–311, 2019.

[13] Rana Nour, Hossam El-Din Moustafa, Ehab H AbdelHay, and Mohamed Maher Ata. Improved unsupervised deep boltzmann learning approach for accurate hand vein recognition. *IEEE Access*, 2024.

[14] Ankit Vidyarthi and Aruna Malik. A hybridized modified densenet deep architecture with clahe algorithm for humpback whale identification and recognition. *Multimedia Tools and Applications*, 81(14):19779–19793, 2022.

[15] Ge-Liang Lv, Lei Shen, Yu-Dong Yao, Hua-Xia Wang, and Guo-Dong Zhao. Feature-level fusion of finger vein and fingerprint based on a single finger image: The use of incompletely closed near-infrared equipment. *Symmetry*, 12(5):709, 2020.

[16] S Surya and A Muthukumaravel. Adaptive sailfish optimization-contrast limited adaptive histogram equalization (asfo-clahe) for hyperparameter tuning in image enhancement. In *Computational Intelligence for Clinical Diagnosis*, pages 57–76. Springer, 2023.

[17] Gabriel Fillipe Centini Campos, Saulo Martiello Mastelini, Gabriel Jonas Aguiar, Rafael Gomes Mantovani, Leonimer Flávio de Melo, and Sylvio Barbon. Machine learning hyperparameter selection for contrast limited adaptive histogram equalization. *EURASIP Journal on Image and Video Processing*, 2019:1–18, 2019.

[18] Prachi Agrawal, Hattan F Abutarboush, Talari Ganesh, and Ali Wagdy Mohamed. Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). *Ieee Access*, 9:26766–26791, 2021.

[19] Selma Trabelsi, Djamel Samai, Fadi Dornaika, Azeddine Benlamoudi, Khaled Bensid, and Abdelmalik Taleb-Ahmed. Efficient palmprint biometric identification systems using deep learning and feature selection methods. *Neural Computing and Applications*, 34(14):12119–12141, 2022.

[20] Anilkumar Gona, M Subramoniam, and R Swarnalatha. Transfer learning convolutional neural network with modified lion optimization for multimodal biometric system. *Computers and Electrical Engineering*, 108:108664, 2023.

[21] Hui-Huang Hsu, Cheng-Wei Hsieh, and Ming-Da Lu. Hybrid feature selection by combining filters and wrappers. *Expert Systems with Applications*, 38(7):8144–8150, 2011.

[22] Ali Dabba, Abdelkamel Tari, Samy Meftali, and Rabah Mokhtari. Gene selection and classification of microarray data method based on mutual information and moth flame algorithm. *Expert Systems with Applications*, 166:114012, 2021.

[23] Majdi M Mafarja and Seyedali Mirjalili. Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing*, 260:302–312, 2017.

[24] Francois Chollet. *Deep learning with Python*. Simon and Schuster, 2021.

[25] Kien Nguyen, Clinton Fookes, Arun Ross, and Sridha Sridharan. Iris recognition with off-the-shelf cnn features: A deep learning perspective. *IEEE Access*, 6:18848–18855, 2017.

[26] Kuo Wang and Ajay Kumar. Toward more accurate

iris recognition using dilated residual features. *IEEE Transactions on Information Forensics and Security*, 14(12):3233–3245, 2019.

[27] Lin Zhang, Zaixi Cheng, Ying Shen, and Dongqing Wang. Palmprint and palmvein recognition based on dcnn and a new large-scale contactless palmvein dataset. *Symmetry*, 10(4):78, 2018.

[28] Eva Tuba, Nebojša Bačanin, Ivana Strumberger, and Milan Tuba. Convolutional neural networks hyperparameters tuning. In *Artificial intelligence: theory and applications*, pages 65–84. Springer, 2021.

[29] Gustavo H De Rosa, João P Papa, and Xin-S Yang. Handling dropout probability estimation in convolution neural networks using meta-heuristics. *Soft Computing*, 22:6147–6156, 2018.

[30] Chia Chin Lip and Dzati Athiar Ramli. Comparative study on feature, score and decision level fusion schemes for robust multibiometric systems. *Frontiers in computer education*, pages 941–948, 2012.

[31] Kyoung Jun Noh, Jiho Choi, Jin Seong Hong, and Kang Ryoung Park. Finger-vein recognition based on densely connected convolutional network using score-level fusion with shape and texture images. *Ieee Access*, 8:96748–96766, 2020.

[32] Mingxing He, Shi-Jinn Horng, Pingzhi Fan, Ray-Shine Run, Rong-Jian Chen, Jui-Lin Lai, Muhammad Khurram Khan, and Kevin Octavius Sentosa. Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recognition*, 43(5):1789–1800, 2010.

[33] Jialiang Peng, Ahmed A Abd El-Latif, Qiong Li, and Xiamu Niu. Multimodal biometric authentication based on score level fusion of finger biometrics. *Optik*, 125(23):6891–6897, 2014.

[34] MMU iris dataset, Available at-. https://www.kaggle.com/datasets/naureenmohammad/mmu-iris-dataset.

[35] CASIA Palmprint Image Database, Available at-. http://www.idealtest.org/dbDetailForUser.do?id=5#/datasetDetail/5.

[36] IIT Delhi Palmprint Image Database version 1.0, Available at-. http://www4.comp.polyu.edu.hk/~csajaykr/IITD/Database_Palm.htm.

[37] Birjand University Mobile Palmprint Database, Available at-. https://www.kaggle.com/datasets/mahdieizadpanah/birjand-university-mobile-palmprint-databasebmpd.

[38] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

[39] Ryan J Urbanowicz, Melissa Meeker, William La Cava, Randal S Olson, and Jason H Moore. Relief-based feature selection: Introduction and review. *Journal of biomedical informatics*, 85:189–203, 2018.

[40] Marko Robnik-Šikonja and Igor Kononenko. Theoretical and empirical analysis of relieff and rrelieff. *Machine learning*, 53:23–69, 2003.

[41] Seyedali Mirjalili. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-based systems*, 89:228–249, 2015.

[42] Jens Behrmann, Will Grathwohl, Ricky TQ Chen, David Duvenaud, and Jörn-Henrik Jacobsen. Invertible residual networks. In *International conference on machine learning*, pages 573–582. PMLR, 2019.

[43] Xin-She Yang. Flower pollination algorithm for global optimization. In *International conference on unconventional computing and natural computation*, pages 240–249. Springer, 2012.

AIP Conference Proceedings

RESEARCH ARTICLE | JULY 24 2024

Mechanical characterization of Al6351 reinforced with TiC, B_4C, and WC: Proportional effects and microstructural analysis FREE

Shahazad Ali ✉; Qasim Murtaza; Pallav Gupta

Check for updates

AIP Conf. Proc. 3111, 060005 (2024)
https://doi.org/10.1063/5.0221454

View Online
Export Citation

06 August 2024 04:48:46

Nanotechnology & Materials Science
Optics & Photonics
Impedance Analysis
Scanning Probe Microscopy
Sensors
Failure Analysis & Semiconductors



Unlock the Full Spectrum.
From DC to 8.5 GHz.
Your Application. Measured.
Find out more
Zurich Instruments

AIP Publishing

Mechanical Characterization of Al6351 Reinforced with TiC, B_4C, and WC: Proportional Effects and Microstructural Analysis

Shahazad Ali[1, a)], Qasim Murtaza[1,b)] and Pallav Gupta[2,c)]

[1]*Mechanical, Production and Industrial Engineering Department, Delhi Technological University, Delhi, 110042, India*
[2]*Department of Mechanical Engineering, Amity School of Engineering and Technology, Amity University Uttar Pradesh, Noida 201313, India.*

Corresponding author: [a)]shahzadali0786@gmail.com, [b)]qasimmurtaza@dce.acin, [c)]pgupta7@amity.edu

Abstract. This paper presents a thorough investigation into the mechanical characterization of Al6351 composites reinforced with TiC, B_4C, and WC particles. The primary objective is to assess how different proportions of reinforcements affect the mechanical properties and microstructure of the composites. The experimental approach involves the fabrication of composite specimens using the liquid route process, followed by comprehensive mechanical and microstructural analyses. The findings demonstrate that increasing the proportions of reinforcements significantly enhances the tensile strength, yield strength, hardness, and toughness of the composites. Microstructural analysis confirms the uniform dispersion of reinforcement particles throughout the matrix alloy, enabling effective load transfer and reinforcement mechanisms. In the hybrid metal matrix composite composed of 1.5 wt.% B_4C, 1.5 wt.% TiC, and 1.5 wt.% WC, the highest tensile strength of 140.7 MPa, and hardness of 76 HRC are observed. Conversely, the hybrid metal matrix composite with 0.5 wt.% of each reinforcement exhibits the maximum toughness of 28 Joules. These research findings offer valuable insights for the design and development of high-performance metal matrix composites with potential applications in industries such as aerospace, automotive, and manufacturing.

06 August 2024 04:48:46

INTRODUCTION

Aluminum alloys encompass a category of metallic substances comprising mainly aluminum as the primary component, accompanied by the inclusion of other elements to augment their characteristics. These alloys present a multitude of benefits in diverse sectors. To begin with, aluminum alloys possess a lightweight quality, rendering them exceptionally suitable for implementations where minimizing weight is crucial, such as within the aerospace, automotive, and transportation fields[1]. Their low density allows for improved fuel efficiency, easier handling, and reduced transportation costs. Additionally, aluminum alloys exhibit excellent corrosion resistance, ensuring durability and longevity in challenging environments[2], [3]. This property eliminates the need for extensive protective coatings and reduces maintenance costs. Another advantage of aluminum alloys is their high formability, allowing for intricate shaping and complex designs through processes like casting, extrusion, and forging[4]. This versatility enables the production of customized components in industries such as construction, consumer goods, and electronics. Furthermore, aluminum alloys have good thermal and electrical conductivity, making them valuable for heat transfer applications and electrical conductivity requirements in industries ranging from heat exchangers to electrical transmission lines[5][6]. Lastly, aluminum alloys are highly recyclable, offering environmental benefits and contributing to the principles of sustainability and circular economy. With these advantages, aluminum alloys have become indispensable materials in numerous industrial sectors, providing a balance of performance, cost-effectiveness, and environmental sustainability[7][8][9].

In the realm of aluminum alloys, specifically Al6351, notable characteristics including formability, robust strength, and corrosion resistance have been identified [10]. However, researchers have sought to amplify its mechanical properties and broaden its range of potential applications by incorporating reinforcing elements,

AICTE & DSIR Sponsored International Conference on Smart and Innovative Development in Science, Engineering & Technology 2023
AIP Conf. Proc. 3111, 060005-1–060005-11; https://doi.org/10.1063/5.0221454
Published under an exclusive license by AIP Publishing. 978-0-7354-5008-0/$30.00

particularly ceramic particles, into the alloy structure. Among the various ceramic reinforcements, TiC, B_4C, and WC hold great potential owing to their remarkable hardness, thermal stability, and resistance to wear[11], [12][13].

The research focuses on examining the mechanical characteristics of Al6351 when reinforced with varying ratios of TiC, B_4C, and WC is driven by the increasing interest in metal matrix composites. These composites have gotten a lot of attention because of their superior mechanical properties, which outperform those of traditional materials.[14]. Because of their lightweight nature and exceptional strength-to-weight ratio, aluminum-based composites have been extensively researched and used in industries such as defense, aviation, and automobile [15],[16],[17]. To optimize the material design and fabrication processes, it is critical to understand the effect of different reinforcement ratios on the mechanical properties of the composite[18]. By adjusting the proportion of ceramic reinforcements, a balance can be achieved between strength, hardness, and other mechanical properties, resulting in improved performance for specific applications[19].

The addition of reinforcing elements, namely TiB2 and ZrB2, to the Al6351 alloy, has notable impacts on its properties. These materials play a key role in improving the tensile strength and hardness of the matrix, Consequently, its overall mechanical performance improves. Furthermore, the inclusion of these reinforcements enhances the composite's resistance to wear, making it well-suited for applications involving abrasive contact or sliding[20][21][22]. The thermal stability of the alloy is also enhanced, as these reinforcing materials have high melting points and thermal stability. Moreover, certain reinforcements like Si3N4 and TiO2 can provide corrosion resistance, protecting the alloy from corrosive environments. Lastly, the electrical and thermal conductivity of the composite can be tailored based on the choice of reinforcing materials[23], [24], [25].

The application of agrowaste and industrial waste as a reinforcingagent to Aluminium based alloys offers significant advantages and diverse applications[26]. In the automotive sector, the incorporation of agrowaste and waste materials in metal matrix composites leads to lightweight components, improving fuel efficiency and reducing emissions. The utilization of agrowaste and industrial waste materials not only provides an eco-friendly solution but also enhances the thermal and electromechanical characteristics of the composites, making them attractive to various industries seeking sustainable and high-performance materials[27], [28].

The importance of this research lies in its contribution to existing knowledge about the mechanical behavior of Al6351 composites. Through a systematic investigation of varying proportions of TiC, B_4C, and WC, valuable insights can be gained into the structure-property relationships, deformation mechanisms, and overall performance of the composites. These findings will offer valuable guidance to materials engineers, researchers, and industries involved in the advancement and application of metal matrix composites.

The lightweight nature of these composites, coupled with their enhanced mechanical properties, can lead to improved fuel efficiency in the automotive sector, reduced structural weight in aerospace applications, and increased durability and wear resistance in defense applications. Additionally, the study contributes to sustainable materials development by utilizing aluminum alloys and ceramic reinforcements, which are abundant and recyclable.

Despite extensive research conducted on metal matrix composites (MMCs), there remains a gap in the literature regarding the mechanical characterization of Al6351 when reinforced with different ratios of TiC, B_4C, and WC. This research gap arises from the necessity to comprehend the precise influence of varying reinforcement compositions on the microstructuralstructure and mechanical properties of hybrid metal matrix composites. The study aims to contribute to the existing knowledge base and provide valuable insights for the advancement and application of advanced MMCs by bridging this research gap.

OBJECTIVES OF THE RESEARCH

The research aims to systematically study the influence of different wt.% proportions of TiC, B_4C, and WC on the mechanical characteristics of Al6351hybrid MMCs. The research endeavors to analyze and compare the outcomes obtained from various reinforcement compositions to determine the ideal proportion that leads to the paramountadvance in mechanical properties.

The research also aims to investigate the microstructure of the Al6351 composites at different reinforcement proportions. Optical microscopic examinations are performed to observe the distribution, size, and morphology of the reinforcing particles within the hybrid MMCs and identify any changes in the microstructure due to varying reinforcement proportions.

06 August 2024 04:48:46

MATERIALS AND PROPERTIES

Aluminium Metal Matrix

Al6351 belongs to the 6xxx series of aluminum alloys and is predominantly comprised of aluminum and magnesium. Widely utilized in engineering applications, Al6351 is a heat-treatable wrought alloy consisting primarily of aluminum (Al), with small additions of magnesium (Mg), silicon (Si), and trace elements. This alloy exhibits a favorable blend of tensile strength, hardness, impact strength, formability, and corrosion resistance, rendering it appropriate for a range of structural and mechanical applications. Its versatility is further enhanced by the ease with which it can be processed using various manufacturing techniques such as casting, extrusion, and rolling, enabling the production of intricate shapes and components.

Reinforcement Materials

Within the realm of hybrid MMCs, the incorporation of reinforcing materials holds significant importance in augmenting the mechanical properties and overall performance of the base metal. Among the various choices available, B_4C,TiC, andWC are extensively utilized as reinforcing materials due to their exceptional characteristics. The specific attributes of these reinforcement materials are elaborated below:

Titanium Carbide (TiC)

Titanium carbide is a refractory ceramic known for its exceptional hardness, outstanding surface scratching resistance, high melting point, and commendable chemical stability. It exhibits a remarkable level of hardness, comparable to that of diamond, thereby significantly enhancing the scratching resistance and tribological aspects of the MMCs and HMMCs. Moreover, TiC possesses a low density, contributing to the overall lightweight nature of the composite and preserving its inherent weight advantages.

Boron Carbide (B_4C)

B_4C is an extremely hard ceramic material known for its exceptional hardness, high melting point, and low density. It exhibits excellent wear resistance, making it suitable for applications where abrasive wear is a concern. is also characterized by its high strength, chemical stability, and neutron absorption capability, making it valuable in defense and nuclear industries.

Tungsten Carbide (WC)

WC is a hard and brittle ceramic material renowned for its exceptional hardness, high melting point, and excellent tribological properties. It exhibits remarkable strength, toughness, and thermal stability, making it suitable for applications requiring high-performance wear resistance. WC particles effectively reinforce the metal matrix composite and enhance its mechanical properties, particularly hardness and abrasion resistance.

TABLE 1. Physical Characteristics of Reinforcement

Reinforcement	Weight density(gram/cc)	Particle size (micron)	Color
Boron Carbide (B_4C)	2.52	80-120	Gray black
Tungsten Carbide (WC)	15.7	40-80	Gray
Titanium Carbide (TiC)	4.91	80-120	Silver gray

06 August 2024 04:48:46

TABLE 2. Sample Nomenclature of AlGr hybrid MMCs

S.N	Al6351	B_4C	TiC	WC	Sample Code
1	100.00	0	0	0	ABTW01
2	98.5	0.50	0.50	0.50	ABTW02
3	97.0	1.00	1.00	1.00	ABTW03
4	95.5	1.50	1.50	1.50	ABTW04
5	94.0	2.00	2.00	2.00	ABTW05

TiC, B_4C, and WC are highly effective reinforcing materials used in metal matrix composites. Their exceptional scratching resistance, tribological properties, strength, and thermal stability contribute to the improved mechanical properties and performance of the composites, making them useful in a variety of industries including automotive, aerospace, defense, and manufacturing.

PREPARATION OF THE HYBRID MMC

The liquid routeprocess is exercised to fabricate MMCs with improved mechanical properties. Desired proportions of TiC, B_4C, and WC particles were selected based on the research objectives and reinforcement characteristics as shown in Table 1.

The process commenced by placing small pieces of Al6351 alloy into a crucible, which was then positioned inside a furnace at a temperature of 925°C, guaranteeing complete liquefaction of the metal. Throughout the melting process, a controlled atmosphere was maintained to prevent oxidation or contamination of the matrix alloy. Subsequently, the metal matrix was heated at 925°C for 30 minutes. Following this, the furnace temperature increased to 970°C. At this stage, preheated and pre-weighed particles of TiC, B_4C, and WC were introduced into the molten metal matrix. To ensure a uniform dispersion and spreading of the reinforcement within the matrix, the mixture underwent intense stirring using a mechanical stirrer. This stirring process played a critical role in breaking up any clumps or agglomerates, promoting effective bonding between the reinforcement and the matrix, and ultimately achieving a homogeneous composite mixture.

To ensure a thorough blending of the reinforcement particles with the molten alloy, the stirring process was continued for 30 minutes. Throughout this process, consistent stirring speed and direction were maintained to minimize the introduction of air bubbles or excessive shear forces that could compromise the integrity of the composite. After 30 minutes, the mixture was left in the furnace for an additional 20 minutes to allow it to settle down.

To prepare the molds, they were preheated at 410°C. This preheating step aimed to reduce the temperature gradient and prevent premature solidification of the molten composite mixture upon contact. The carefully controlled molten mixture of the composite was then poured into the molds, taking care to avoid splashing or introducing impurities. The molds were then allowed to cool down, resulting in the solidification of the molten mixture into a solid form. The solidification process took place in a still-air environment. Once the molds had cooled down completely, the fabricated hybrid MMCs were removed from the molds. Post-processing steps, including trimming excess material and removing surface defects, were carried out on the cast specimens.

This entire process was repeated for five different compositions to fabricate hybrid metal matrix composites. The resulting samples were labeled according to their composition, as indicated in Table 2.

EXPERIMENTATIONS

In the study centered on investigating the mechanical characteristics of Al6351 alloy reinforced with TiC, B_4C, and WC, a range of testing equipment and methodologies are utilized to assess the mechanical properties and analyze the microstructure of the composite specimens. The specific tests performed are elaborated upon below:

06 August 2024 04:48:46

Microstructural Analysis

Microstructural analysis techniques are used to examine the internal structure, distribution, and morphology of the reinforcement particles within the hybrid MMCs.The sample preparation procedure for scanning electron microscope testing involves several steps to ensure the proper examination of the material's microstructure. Firstly, a representative sample was selected from the material of interest. This sample was then cut into a smaller, manageable size using cutting tools or precision cutting machines. The next step was mounting, where the specimen was embedded in a mounting medium to provide support and stability. Grinding was performed using abrasive papers or grinding wheels to eliminate surface irregularities and scratches. Subsequently, polishing was conducted using progressively finer nylon cloth to achieve a smooth and mirror-like surface. For optical microscopy, the prepared samples were mounted on a microscope and the micrographs were taken using the embedded camera. Micrographs of the specimen were captured using microscopytechniques are shown in Figure 1.

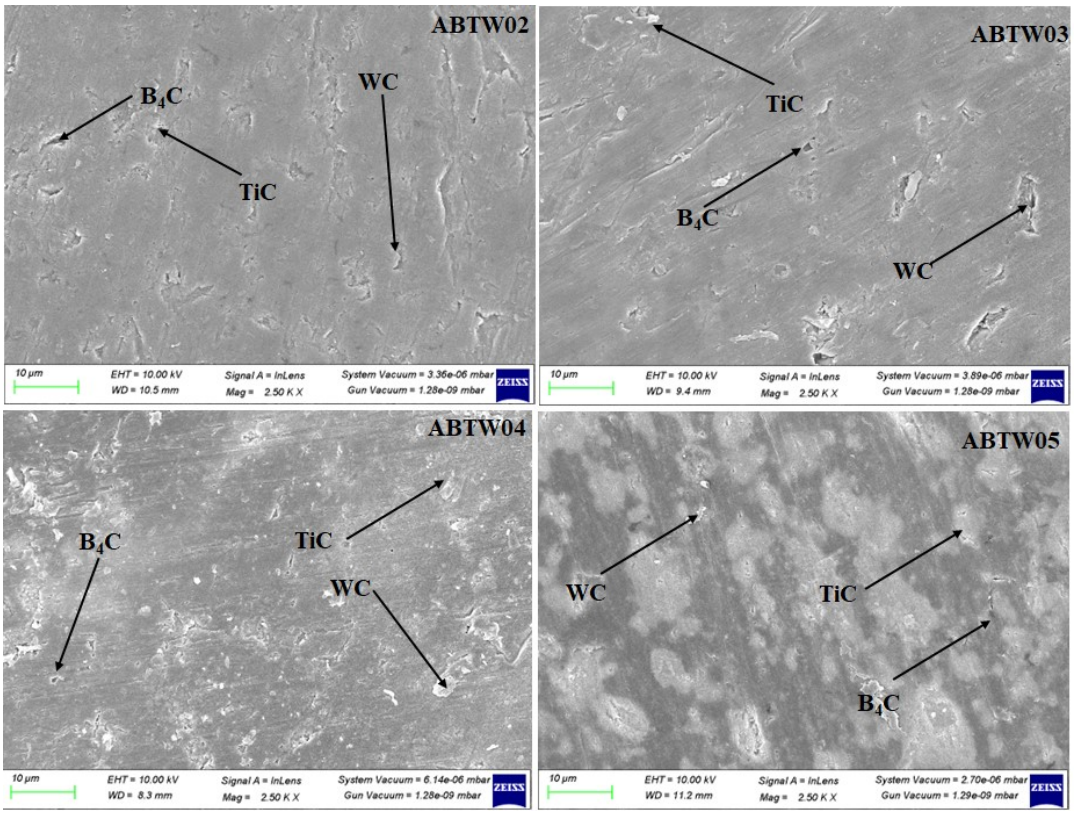


FIGURE 1: Optical Micrograph of Fabricated HMMCs

The manufactured hybrid metal matrix composite incorporates a small amount of lightweight reinforcing particles, which are uniformly dispersed within the metal matrix. This composite exhibits a consistent dispersion of the reinforcing particles and showcases robust intermetallic bonding.Sample ABTW02, AWTW03, and ABTW04 show a uniform dispersion of reinforced particles. However, when the weight percentage of reinforced particles surpasses 1.5%, clustering of the particles occurs, resulting in compromised intermetallic bonding. Sample ABTW05 with 2 wt.% of each reinforcement, shows huge clustering of the particles.

Mechanical Testing:

Hardness Testing:

Hardness tests are conducted to assess the hardness properties of the composite specimens, which provide insights into their resistance to localized deformation or indentation.Common hardness testing methods include

06 August 2024 04:48:46

Vickers, Rockwell, or Brinell hardness tests. A specimen of 10 x10 x 10 mm was fabricated to measure the Rockwell hardness of the specimen. Five hardness measurements were taken on each specimen. The hardness of Fabricated samples is shown graphically in Figure 2.

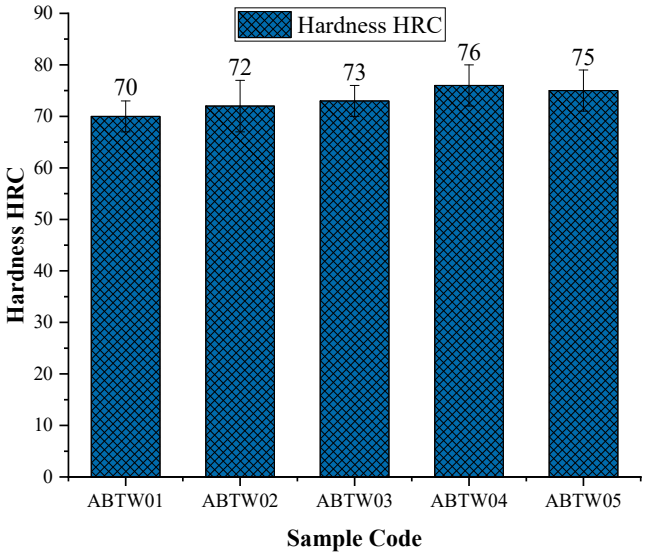


FIGURE 2: Hardness of Fabricated Specimen

Impact Testing

The impact test serves as a vital mechanical evaluation to assess the resilience and ability of a material to withstand abrupt loading. This test involves subjecting a standardized specimen to a rapid and high-velocity load, usually achieved by striking the specimen with a pendulum or hammer in a standardized manner. It provides valuable insights into the material's toughness and resistance to sudden loading. The impact test results provide valuable information about the toughness and fracture behavior of Al6351 alloy. It helps assess the material's ability to withstand sudden loading conditions, such as impact or collision, which is crucial in industries where the alloy is used for structural or safety-critical components.Fabricated specimens as per ASTM standards are shown in Figure 3.

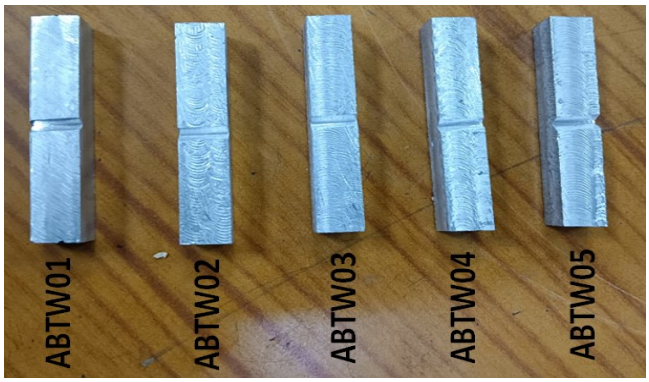


FIGURE 3: Charpy Test specimen

To perform the impact test, standardized Charpy specimens of hybrid MMCs were prepared using a slotter tool. The specimens were then placed in an impact testing machine as a simply supported beam. The machine delivered a high-velocity jolt to the sample, triggering it to crack. The amount of energy absorbed by the specimen was recorded from the scale. Three specimens of each composition were tested to obtain average impact strength. The average impact strength of each composition with experimental error is shown in Figure 4.

06 August 2024 04:48:46

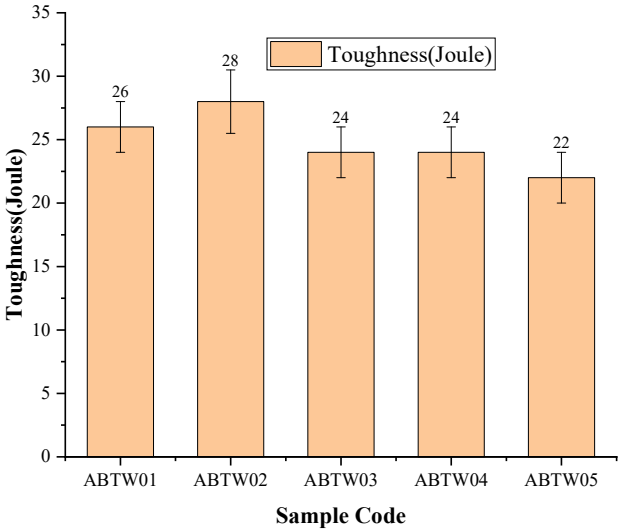


FIGURE 4: Toughness of Fabricated Specimen

Tensile Testing

After the post-processing of cast samples, three specimens for each composition were prepared for tensile testing. Tensile tests were conducted on a PC2000 tensiometer having a load capacity of 20kN. Testing samples were prepared as ASTM E8M standard. Figure 1 depicts tensiometer samples of Al6351 and hybrid MMCs.

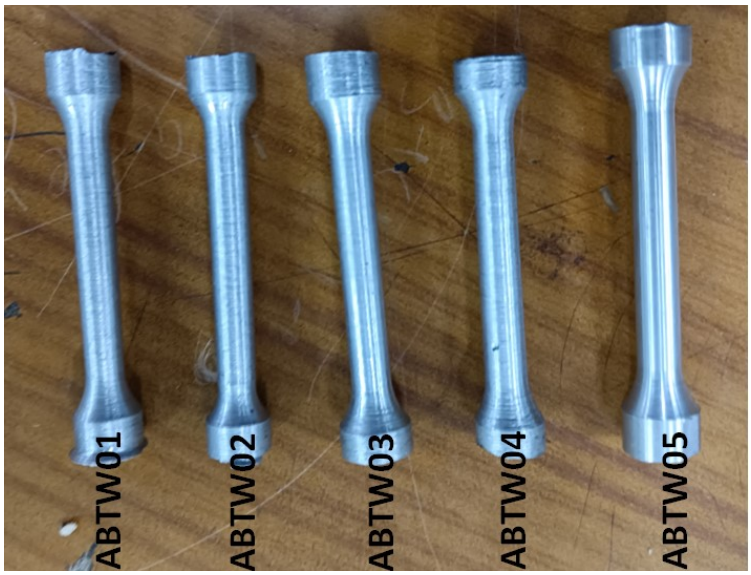


FIGURE 5: Tensiometer Specimen

The test report includes the measurements of Engineering UTS, True UTS, and elongation. The recorded results are presented in Table 3.

TABLE 3: Stress-Strain data of Hybrid Metal Matrix Composite

Sample	Eng UTS(MPa)	True UTS(MPa)	Break Displacement(mm)
ABTW01	100.2	136.7	3.92
ABTW02	117.5	141.8	2.93

06 August 2024 04:48:46

ABTW03	131.7	158.9	2.88
ABTW04	140.7	168.9	2.08
ABTW05	127.9	155.5	1.96

DISCUSSION

The microstructure of composites portrays a pivotal role in concluding mechanical properties. Uniform distribution of ceramic particles within the metal matrix facilitates consistent reinforcement and minimizing localized stress concentrations, effective load transfer, and enhances overall mechanical performance.

The bonding between the reinforcement particles and the matrix also plays animportant role. A strong and well-bonded interface enables efficient load transmission and improves the mechanical properties of the composite.

The microphotograph clearly shows that at lower reinforcement proportions, there is a uniform dispersion and distribution of the reinforcement particles, along with strong interfacial bonding. However, at higher reinforcement proportions, the distribution of reinforcement particles becomes non-uniform, leading to clustering and weak bonding between the reinforcement and the matrix metal.

Initially, the addition of B_4C, TiC, and WC reinforcing elements leads to an increase in hardness. This is because these materials typically have higher hardness than the Al6351 alloy. When incorporated into the matrix, they act as strengthening agents, hindering the movement of dislocations and impeding plastic deformation. As a result, the composite exhibits improved resistance to deformation, leading to higher hardness values.However, as the concentration of reinforcing elements increases beyond 1.5 wt.% each, the hardness start to decrease. This is attributed to the formation of clustering or agglomeration of the reinforcing particles, which create regions of stress concentration and lead to localized plastic deformation. These agglomerations act as initiation points for crack formation, reducing the overall hardness of the composite.

The inclusion of B_4C, TiC, and WC reinforcing elements in the composite leads to a slight improvement in impact strength. These reinforcing particles serve as obstacles, hindering crack propagation and dissipating energy during impact loading. The increased strength and hardness resulting from these reinforcements contribute to enhanced impact resistance. However, when the concentration of reinforcing elements exceeds 0.5 wt.% of each, the impact strength may diminish. Several factors contribute to the decrease in impact strength of the hybrid MMCs. One primary reason is the introduction of brittleness into the composite by adding reinforcement. Another reason is the potential formation of agglomerations or clusters of reinforcing particles, resulting in regions of stress concentration that initiate crack propagation. This clustering effect compromises the overall impact resistance of the composite. Furthermore, weak interfacial bonding between the reinforcing particles and the matrix material can hinder load transfer and impede energy absorption during impact loading. Insufficient interfacial bonding further contributes to the reduction in impact strength.

The mechanical testing results indicated a substantial enhancement in the tensile strength of the hybrid MMCs as the proportion of reinforcement increased. The hybrid MMCs with higher reinforcement proportions exhibited superior mechanical properties in comparison to the unreinforced Al6351 alloy. These improvements in mechanical properties can be ascribed to the strengthening effect caused by the reinforcement particles, which impede dislocation movement and enhance load transfer within the composites. The composite sample ABTW03, containing 1.5 wt.% B_4C, 1.5 wt.% TiC, and 1.5 wt.% WC demonstrated the highest engineering UTS among all the fabricated samples. However, further increasing the weight-wisepercentage of reinforcement led to a decrease in tensile strength. Table 3 indicates a continuous decrease in break displacement, indicating an increase in the brittleness of the hybrid metal matrix composite. Figure 6 graphically represents the Engineering UTS and True UTS.

The decrease in strength observed can be attributed to the agglomeration of reinforcement particles. Specimens with lower proportions of reinforcement exhibited a more uniform distribution and better bonding of reinforcement within the metal matrix. The results clearly indicated a correlation between the microstructure and mechanical properties. The presence of a uniform distribution of reinforcement particles and a strong interface between the particles and the matrix contributed to improved mechanical properties. However, when comparing these findings to previous studies, the addition of a third reinforcement instead of two did not demonstrate a significant impact on the tensile strength of the hybrid metal matrix composite[29].

06 August 2024 04:48:46

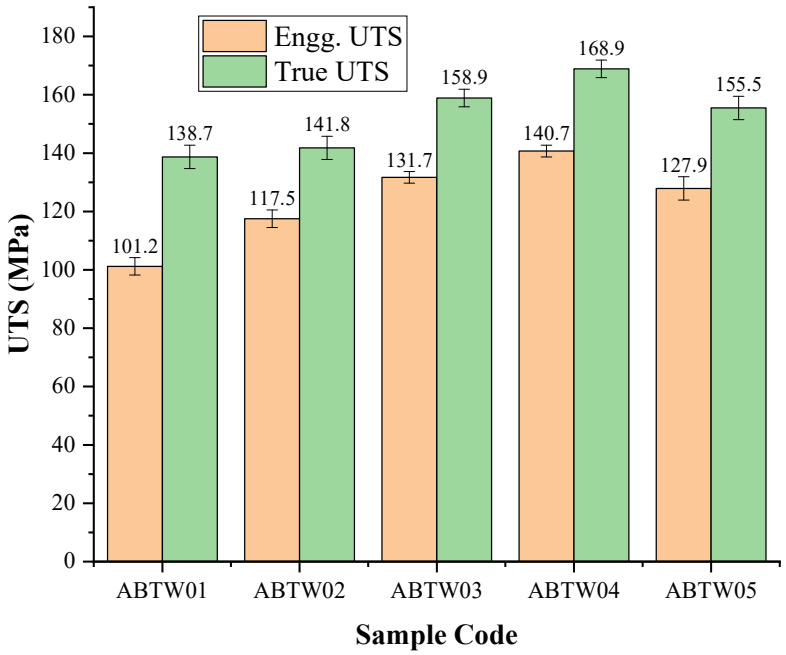


FIGURE6: Engg. UTS and True UTS of Fabricated Specimen

CONCLUSION

The research primarily focused on investigating the mechanical characteristics of Al6351 alloy reinforced with TiC, B_4C, and WC. The study yielded several noteworthy findings, including the following:

i. The microphotograph analysis confirmed the presence of a uniform distribution of reinforcement particles within the hybrid MMCs.
ii. The interface between the particles and the matrix demonstrated strong bonding, indicating effective load transfer between the reinforcement and matrix.
iii. The incorporation of TiC, B_4C, and WC reinforcement particles into the Al6351 alloy led to significant enhancements in mechanical properties.
iv. The ultimate tensile strength and hardness of the composite specimens showed a progressive increase with higher proportions of reinforcement.
v. The Hybrid MMC sample containing 1.5 wt. % of each reinforcement exhibited the utmostengineering UTS of 140.7 MPa and hardness of 76 HRC.Maximum toughness was observed in Hybrid MMC having 0.5 wt.% of each reinforcement as 28 Joule.
vi. However, beyond a certain reinforcement proportion, there was a diminishing return in the enhancement of properties, suggesting an optimal reinforcement range.

In brief, the research findings indicate that increasing the reinforcement proportions of TiC, B_4C, and WC in Al6351 alloy leads to improved mechanical properties and enhanced wear resistance. However, there is an optimal range of reinforcement proportions beyond which the benefits become less significant.

REFERENCES

1. Wu, Y. (2022, March). Application of aluminum alloy in aircraft. In Journal of Physics: Conference Series (Vol. 2228, No. 1, p. 012024). IOP Publishing.
2. Ahamad, N., Mohammad, A., & Gupta, P. (2021). Wear characteristics of Al matrix reinforced with Al2O3-carbon hybrid metal matrix composites. Materials Today: Proceedings, 38, 63-68.
3. Ahamad, N., Mohammad, A., & Gupta, P. (2021). Wear characteristics of Al matrix reinforced with Al2O3-carbon hybrid metal matrix composites. *Materials Today: Proceedings*, *38*, 63-68..

06 August 2024 04:48:46

4. Toros, S., Ozturk, F., & Kacar, I. (2008). Review of warm forming of aluminum–magnesium alloys. *Journal of materials processing technology*, *207*(1-3), 1-12.
5. E. Georgantzia, M. Gkantou, and G. S. Kamaris, "Aluminium alloys as structural material: A review of research," *Eng Struct*, vol. 227, p. 111372, Jan. 2021, doi: 10.1016/J.ENGSTRUCT.2020.111372.
6. Qu, X. H., Zhang, L., Mao, W. U., & Ren, S. B. (2011). Review of metal matrix composites with high thermal conductivity for thermal management applications. *Progress in Natural Science: Materials International*, *21*(3), 189-197.
7. Capuzzi, S., &Timelli, G. (2018). Preparation and melting of scrap in aluminum recycling: A review. *Metals*, *8*(4), 249.
8. Jirang, C. U. I., & Roven, H. J. (2010). Recycling of automotive aluminum. *Transactions of Nonferrous Metals Society of China*, *20*(11), 2057-2063..
9. Moghimian, P., Poirié, T., Habibnejad-Korayem, M., Zavala, J. A., Kroeger, J., Marion, F., & Larouche, F. (2021). Metal powders in additive manufacturing: A review on reusability and recyclability of common titanium, nickel and aluminum alloys. *Additive Manufacturing*, *43*, 102017.
10. Kumar, S., Kumar, D., Jain, J., & Hirwani, J. K. (2016). Influence of load, sliding speed, and microstructure on wear response of AZ91 Mg alloy. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *230*(12), 1462-1469..
11. Shorowordi, K. M., Laoui, T., Haseeb, A. A., Celis, J. P., & Froyen, L. (2003). Microstructure and interface characteristics of B_4C, SiC and Al2O3 reinforced Al matrix composites: a comparative study. *Journal of Materials Processing Technology*, *142*(3), 738-743..
12. Veeravalli, R. R., Nallu, R., & Mohiuddin, S. M. M. (2016). Mechanical and tribological properties of AA7075–TiC metal matrix composites under heat treated (T6) and cast conditions. *Journal of Materials research and Technology*, *5*(4), 377-383..
13. Ghasali, E., Alizadeh, M., &Ebadzadeh, T. (2015). hosseinPakseresht A., Rahbari A., Investigation on microstructural and mechanical properties of B_4C–aluminum matrix composites prepared by microwave sintering. *J. Mater. Res. Technol*, *4*(4), 411-415..
14. Sun, J., Zhao, J., Gong, F., Ni, X., & Li, Z. (2019). Development and application of WC-based alloys bonded with alternative binder phase. *Critical Reviews in Solid State and Materials Sciences*, *44*(3), 211-238..
15. Wang, G., Zhao, Y., & Hao, Y. (2018). Friction stir welding of high-strength aerospace aluminum alloy and application in rocket tank manufacturing. *Journal of Materials Science & Technology*, *34*(1), 73-91.
16. Ahamad, N., Mohammad, A., Sadasivuni, K. K., & Gupta, P. (2020). Structural and mechanical characterization of stir cast Al–Al2O3–TiO2 hybrid metal matrix composites. *Journal of Composite Materials*, *54*(21), 2985-2997..
17. Ahamad, N., Mohammad, A., Sadasivuni, K. K., & Gupta, P. (2020). Phase, microstructure and tensile strength of Al–Al2O3–C hybrid metal matrix composites. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *234*(13), 2681-2693.
18. Nadda, R., Kumar, R., Singh, T., Chauhan, R., Patnaik, A., &Gangil, B. (2018). Experimental investigation and optimization of cobalt bonded tungsten carbide composite by hybrid AHP-TOPSIS approach. *Alexandria engineering journal*, *57*(4), 3419-3428.
19. Nwobi-Okoye, C. C., & Uzochukwu, C. U. (2020). RSM and ANN modeling for production of Al 6351/egg shell reinforced composite: Multi objective optimization using genetic algorithm. *Materials Today Communications*, *22*, 100674..
20. Mohanavel, V., Rajan, K., Arul, S., & Senthil, P. V. (2017). Production, Microstructure and Mechanical behavior of AA6351/TiB2 composite synthesized by direct melt reaction method. *Materials Today: Proceedings*, *4*(2), 3315-3324..
21. Mohanavel, V., Ravichandran, M., & Kumar, S. S. (2020). Tribological and mechanical properties of Zirconium Di-boride (ZrB2) particles reinforced aluminium matrix composites. *Materials Today: Proceedings*, *21*, 862-864.
22. Mohanavel, V., Rajan, K., & Senthil Kumar, K. R. (2015). Study on mechanical properties of AA6351 alloy reinforced with titanium di-boride (TiB2) composite by in situ casting method. In *Applied Mechanics and Materials* (Vol. 787, pp. 583-587). Trans Tech Publications Ltd.
23. Zenou, V. Y., & Bakardjieva, S. (2018). Microstructural analysis of undoped and moderately Sc-doped TiO2 anatase nanoparticles using Scherrer equation and Debye function analysis. *Materials Characterization*, *144*, 287-296..

06 August 2024 04:48:46

24. Sundram, S., Raman, M. S., &Balamuralitharan, S. (2023). Influence of process parameters on machining studies on stir casted MMCs with AA6351 and TiO2 by grey and desirability approaches. *Materials Today: Proceedings*, *77*, 551-556..
25. Mohanavel, V., Ali, K. A., Prasath, S., Sathish, T., & Ravichandran, M. (2020). Microstructural and tribological characteristics of AA6351/Si3N4 composites manufactured by stir casting. Journal of Materials Research and Technology, 9(6), 14662-14672.
26. Arora, G., & Sharma, S. (2018). A comparative study of AA6351 mono-composites reinforced with synthetic and agro waste reinforcement. *International Journal of Precision Engineering and Manufacturing*, *19*(4), 631-638.
27. Arunkumar, S., & Kumar, A. S. (2021). Studies on egg shell and SiC reinforced hybrid metal matrix composite for tribological applications. Silicon, 1-9.
28. Sydow, Z., Sydow, M., Wojciechowski, Ł., & Bieńczak, K. (2021). Tribological performance of composites reinforced with the agricultural, industrial and post-consumer wastes: A review. Materials, 14(8), 1863.
29. Ali, S., Murtaza, Q., & Gupta, P. (2023). Synergetic Effect of Gr-B_4C Reinforcement on the Structural and Mechanical Properties of AA6351 Hybrid Metal Matrix Composites. ECS Journal of Solid State Science and Technology.

06 August 2024 04:48:46

2024 International Conference on Computational Intelligence and Computing Applications (ICCICA) | 979-8-3503-0644-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCICA60014.2024.10584930

Motor Imagery Features Selection with Grey Wolf Optimization for Brain Computer Interface

Pawan
Department of Computer Science and Engineering, Meerut Institute of Engineering Technology, Meerut, Uttar Pradesh, India
pawan.panchal@miet.ac.in

Rohtash Dhiman
Department of Electrical Engineering Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana, India
rohtash.ee@drcrustm.org

Jasbir Singh Saini
Director of Panipat Institute of Engineering and Technology
Panipat Institute of Engineering and Technology, Samalkha, Haryana, India
jssain@rediffmail.com

Ankur Saxena
Department of Computer Science and Engineering
Meerut Institute of Engineering Technology, Meerut, Uttar Pradesh, India
ankursaxena19june@gmail.com

Vikas Sharma
Department of Information Technology
Delhi Technological University, Delhi, India
vikassharma12387@gmail.com

***Abstract*— Feature selection is a significant process in brain-computer interfaces (BCIs) that has the potential to significantly enhance system performance, especially if a large number of features are present. Feature selection can help decrease noisy data, overtraining effects, storage needs, and computational complexity thereby enhancing classifier performance. The various feature selection strategies have been reported to attain these objectives. In this work, a hybrid feature selection technique is suggested. It uses a wavelet packet decomposition technique to extract features and grey wolf optimization (GWO) technique to obtain the best subset of features. It is then evaluated with support vector machine (SVM) classifier and obtains maximum classification accuracy of 82.88% for the approximate entropy feature. Experimental findings reveal that the suggested approach outperforms the contemporary MI-BCI systems even with relatively small training data sets.**

Keywords*— Brain computer interfaces (BCIs), Feature extraction, Grey Wolf Optimization (GWO), Electroencephalography (EEG), Wavelet Packet Decomposition (WPD)

I. INTRODUCTION

The EEG, initially recorded for individuals by Hans Berger in 1929, provides information on neural activity in mind. It is a powerful approach for investigating the brain's thinking processes. [1]. Surface electroencephalography is a non-invasive approach that allows recording the biological electrical movement in individual brain, which is vital for various practical implications to improve life of a person suffering from locomotor difficulties [2]. EEG recordings are crucial in BCI applications for sensorimotor functions and established communication in patients with severe motor disabilities [3]. The applications of motor imagery BCI also extend to healthcare, navigation, military services, virtual gaming, robotics, communication and controls. Analysis of EEG signals is required for designing of BCI system. BCIs refer to the mutual interaction between human brains and computers, which can be examined via EEG-based motor imagery signals [4]. EEG-captured motor imagery (MI) signals are the most realistic foundation for BCI development and MI can affect neuronal activity in the subject brain. To control the device by BCI, it is essential to design an algorithm of channel selection, feature extraction, and selection associated to sensorimotor area of human brain [5]. Manual analysis of these processes by experts is time-consuming and costly. Nowadays, many techniques are appearing in literature to automate the analysis of EEG signals, and EEG-based MI BCI system provides a significant degree of mobility, which enables disable people to operate device. The sensorimotor region of individual brain-related feature selection and classification techniques significantly influences the success of the MI-BCI system [6]. Several intelligent and analytical methods based upon wavelet transform (WT), support vector machine (SVM), neural networks (NNs) [7] , evolutionary computation, principal component analysis (PCA), Genetic algorithms (GA) [8], correlation coefficient, grey wolf optimization (GWO), k-nearest neighbors (k-NN) and many more for analysis, classification, and implementation of a control system using MI signals are appearing in the literature [9] [10]. Although many intelligent methods appear in the literature for extracting knowledge and classification from EEG signals, still a lot of research work is needed to extract useful information from these signals to develop an effective BCI system that can help in improving human lives in terms of the design of perfect prosthetics for disabled subjects. A great challenge is to design an effective classification mechanism for EEG signals for different motor imagery activities for BCI systems. Although SVM is the most common classifier and offers numerous advantages, like noise tolerance and adaptivity, more research needs to be done to combine SVM with GWO for feature selection. The outcomes of the studies show that the proposed work effectively detects MI activity using EEG data. The proposed work can be used for seriously disabled people, mind reading, emotion finding and entertainment. The rest of paper is arranged as follows: Section 2 describes materials & methods. The findings and discussion are

Copyright © IEEE–2024 ISBN: 979-8-3503-0644-6

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:46 UTC from IEEE Xplore. Restrictions apply.

explained in Section 3. The section 4 finally, presents the conclusion with future scope.

Contribution of the present work are summarized:

- Perform WPD up to the 8th level for the 8 to 32 Hz frequency range on the motor imagery electroencephalogram (MI-EEG) data.
- Mean, variance, standard deviation, and approximate entropy features are computed separately for 124 WPD nodes out of 256. These nodes are significantly associated with motor imager activities.
- GWO technique is used to selects optimal features for the MI signals.
- Benchmarked machine learning algorithm SVM is used for the classification process.

II. METRIAL AND METHODS USED

A. Dataset

The open-access "BCI Competition IV-dataset I" is employed in this work to analyse MI signals [11]. This dataset includes information from 7 participants and contains fifteen-nine channels sampled at 1000 Hz for motor imagery activities and participant chose two motor imagery activities out of three. To save computing work, 100 Hz down-sampled data was utilized in the present study, with each healthy person executing 200 trials. Participants visualized a single motor imagery activity in each trial for 4 seconds. In the timing scheme depicted [10], the subjects were given a stimulus from which the beginning and end times of an M.I. task could be precisely specified. Every participant provided a cue according to timing method, ensuring that the starting and end times for an MI activity were accurate. Table 1 shows the dataset description for proposed study.

TABLE I. DATASETS FOR FOUR HEALTHY SUBJECTS

Subjects	Data size	Motor imagery activities
a	$190549 \times 59\ int16$	Left hand and foot
b	$190549 \times 59\ int16$	Right and left hand
f	$190608 \times 59\ int16$	Left hand and foot
g	$190602 \times 59\ int16$	Right and left hand

III. FEATURE EXTRACTION TECHNIQUE

For EEG processing, feature extraction plays a vital role. Wavelet Transform (WT) is a generally employed for extracting features in time-frequency domain for BCIs. The wavelet packet decomposition (WPD) allows for a various-level time-frequency decomposition for input EEG signals [12]. Wavelet transform can represent any joint function as an endless sequence of wavelets. The suggested study uses WPD at the 8th level, generating 256 nodes. EEG data with frequencies ranging from 8 to 32 Hz include MI activities. For the 8-32 Hz frequency region, 124 out of 256 wavelet nodes of 14 channels are used to compute the mean, standard deviation, variance and approximate entropy (Table 2) feature vectors size 14×124 separately for sensorimotor related area of subject's brain. The Fig.1 depicts the WPD tree structure for the eighth-level decomposition and fig. 2 shows EEG signal and WPD nodes.

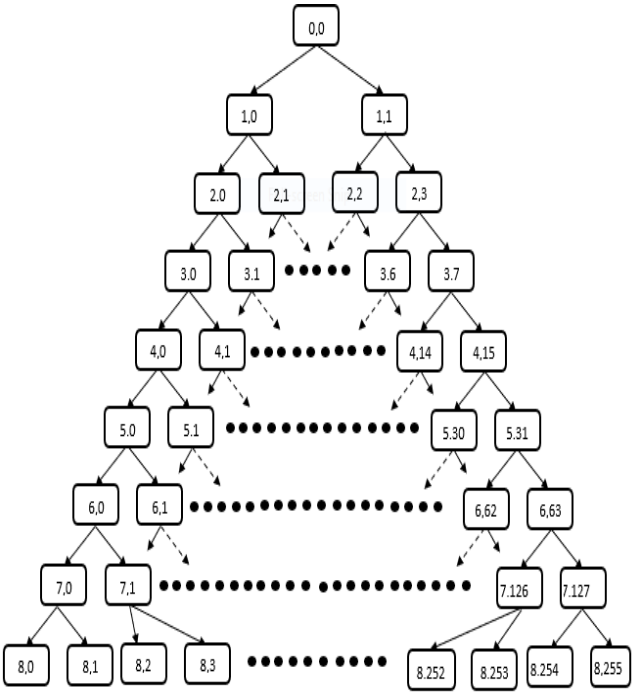


Fig. 1. WPD tree structure up to 8th level [12]

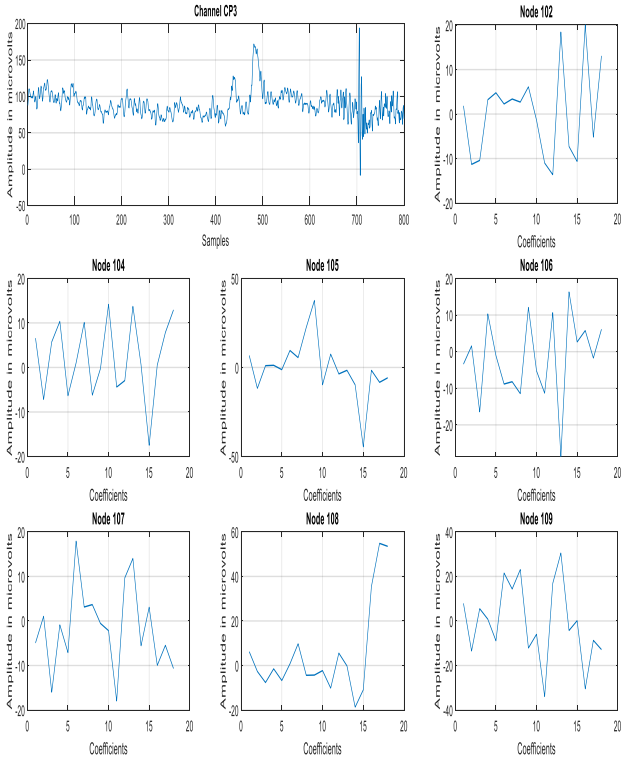


Fig. 2. MATLAB graph of WPD node coefficients (8,102, 8,104 to 8,109) at eighth level

Copyright © IEEE–2024 ISBN: 979-8-3503-0644-6

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:46 UTC from IEEE Xplore. Restrictions apply.

TABLE II. MATHEMATICAL EQUATIONS OF FEATURES EEG [12]

S. No.	Features	Equations
1	Mean (μ)	$\frac{1}{N}\sum_{n=1}^{N} f(n)$
2	Standard deviation (σ)	$\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(f(n)-\mu)^2}$
3	Variance (σ^2)	$\frac{1}{N-1}\sum_{n=1}^{N}(f(n)-\mu)^2$
4	Approximation Entropy (ApEn)	$(r)-\varphi^{m+1}(r)$ [13]

A. Grey Wolf Optimizer (GWO)

Mirjalili was motivated by grey wolves' social behaviour and hunting methods [14]. Grey wolves are generally into found in packs of five to twelve members. The population of GWO is divided generally into four groups**:** $alpha\ (\alpha), delta\ (\delta), beta\ (\beta)$ and $omega\ (\omega)$. The renowned leader of the pack, the alpha wolf, decides which individuals will move locations. As the second leader, the beta wolf helps the alpha decide which actions to take next and takes care of the pack if the alpha is sick or dead. The third-ranking wolf in pack is the delta wolf, responsible for controlling omega wolves. The best three most mathematically fit solutions in GWO are α, β, and δ. The hunting operation in GWO is led by α, β, and δ and ω follows them. The encircling behaviour of a chasing pack is generally denoted as:

$$\vec{P}(t+1) = \overrightarrow{P_p}(t) - \overrightarrow{F_1} \cdot \vec{D} \quad (1)$$

$$\vec{D} = \left|\overrightarrow{F_2} \cdot \overrightarrow{P_p}(t) - \vec{P}\,(t)\right| \quad (2)$$

Here, $\vec{P}$ is position of grey wolf, $\overrightarrow{P_p}$ is position of prey and distance vector between prey and grey wolf is represented by D. The $\overrightarrow{F_1}$ & $\overrightarrow{F_2}$ are coefficient vectors computed with following equation:

$$\overrightarrow{F_1} = 2\vec{a} \cdot \overrightarrow{r_1} - \vec{a} \quad (3)$$

$$\overrightarrow{F_2} = 2 \cdot \overrightarrow{r_2} \quad (4)$$

In this work, r_1 & r_2 are values chosen randomly from [0,1]. According to equation 5, the surrounding coefficient (a) decreases gradually from 2 to 0 over iterations [15] and z_1 and z_2 in fitness function are 0.99 and 0.01 respectively.

$$a = 2 - 2\left(\frac{t}{T}\right) \quad (5)$$

The wolf's new location is updated by equation (6) [16].

$$\vec{P}(t+1) = \frac{\overrightarrow{P_1}+\overrightarrow{P_2}+\overrightarrow{P_3}}{3} \quad (6)$$

Emary et al. [17] introduced GWO techniques to overcome feature selection difficulties.

B. Support Vector Machine (SVM)

SVM machine learning that was initially introduced for the classification of data. The aim is to find the hyperplane that separates training data based on expected risk [18] . SVM offers excellent accuracy for classification due to the maximum margin of hyperplanes [19] It is more flexible than standard neural networks because it includes multiple parameters to get better classification. SVM is seldom influenced by the well-known "curse of dimensionality." A optimal hyperplane obtained by solving the constrained optimisation problem [20]. Optimal hyperplane is achieved by linear SVMs by minimizing the cost function below with constraints.

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \quad (5)$$

$$y_i(w^T x_i + b) \geq 1 - \xi_i \quad (6)$$

Here, $\xi_i \geq 0$, $\|.\|^2$ represents euclidean norm, slack variables expressed by ξ_i, b is the bias, and C is a regularisation parameter. Selecting the correct C value is crucial for balancing complexity and non-separability. Table 3 depicts the parameters used in this proposed work.

TABLE III. PARAMETERS USED FOR SVM

Technique	Hyperparameter	min	max	Step size
SVM *rbf Kernel*	C is kernel coefficient gamma	0.01	0.1	1

IV. RESULTS AND DISCUSSION

The results of classifying motor imagery BCI data for all four participants in this research are present in Table 4 and Figure 4. Each channel is decomposed with WPD at the 8th level, obtaining 256 coefficients. The 124 wavelet packet coefficients representing the 8-32 Hz frequency range with M.I. activity are chosen for further processing. The mean, standard deviation, variance and approximate entropy features are calculated separately for selected nodes. The feature vector is generated for 14 channels (mentioned in section 3.1), spanning the brain's main motor imagery area [6]. The feature vectors have dimensions of 14×72=1008 obtained for four different features using GWO techniques. The suggested approach is tested using ten-fold cross-validation. The Proposed method obtained the maximum classification accuracy of 74.62%, 78.02%, 73.45% and 82.88% for mean, standard deviation, variance, and approximate entropy features respectively. Figure 3 shows proposed work flowchart. The Fig. 4 and Table 4 show classification findings for selected features.

Copyright © IEEE–2024 ISBN: 979-8-3503-0644-6

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:46 UTC from IEEE Xplore. Restrictions apply.

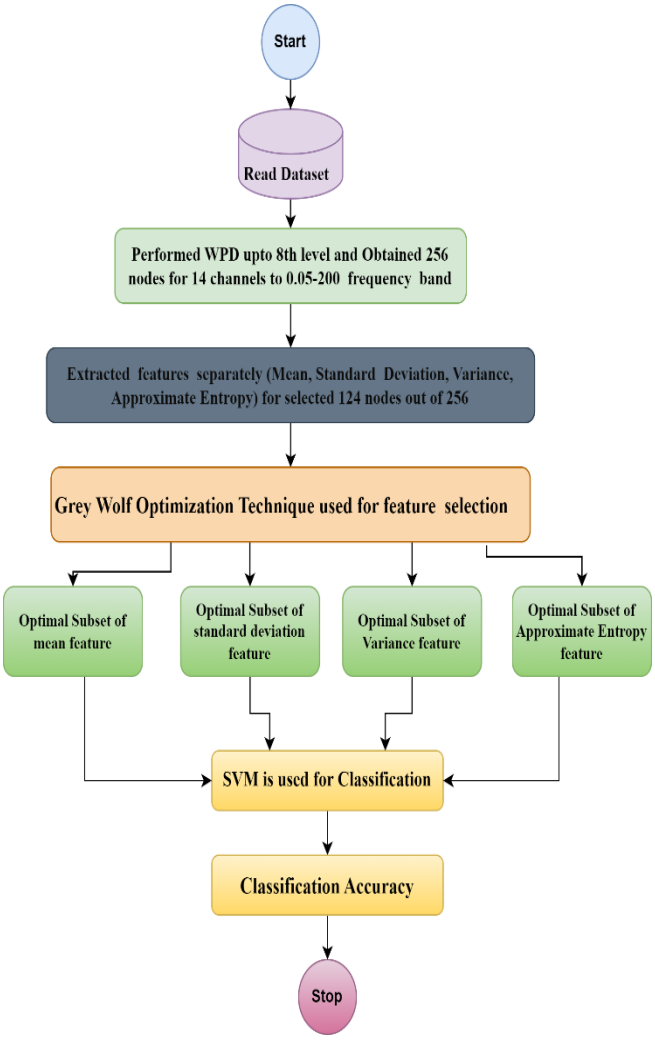


Fig. 3. Froposed work flowchart

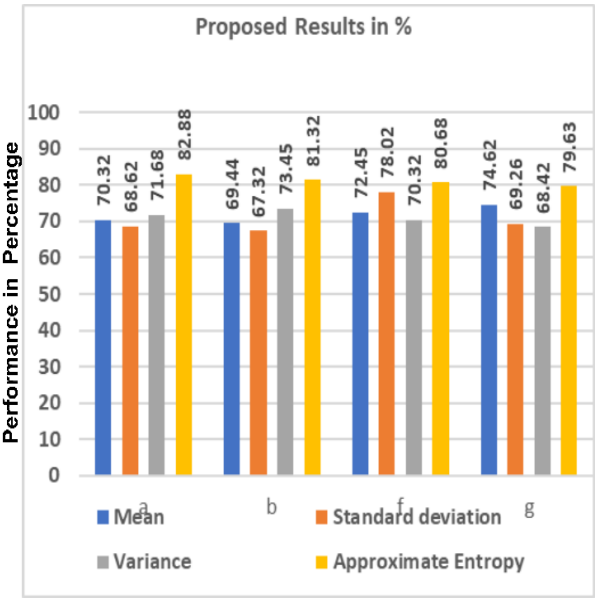


Fig. 4. Results for proposed work

TABLE IV. CLASSIFICATION RESULTS FOR PROPOSED WORK

S.No	Feature	Maximum Classification Accuracy (%)
1	Mean (μ)	74.62
2	Standard deviation (σ)	78.02
3	Variance (σ^2)	73.45

V. CONCLUSION WITH FUTURE SCOPE

The BCI permits injured patients to interact and operate equipment without relying on their peripheral muscles and nerves. The suggested technique is tested using data from four participants (a,b,f,g) in dataset I of BCI-Competition-IV. The suggested research used a WPD approach to extract features (mean, standard deviation, variance, and approximation entropy) and a grey wolf optimisation technique to identify the optimal subset features. It is then analysed using a support vector machine classifier, achieving an optimal classification accuracy of 82.88% for the approximate entropy feature. The method has been developed using MATLAB software. The future objectives involve reducing the dimensionality of feature vectors using different optimization techniques to improve classification accuracy for brain computer interface applications.

REFERENCES

[1] R. Dhiman, J. S. Saini, and Priyanka, "Genetic algorithms tuned expert model for detection of epileptic seizures from EEG signatures," *Appl. Soft Comput. J.*, vol. 19, pp. 8–17, 2014, doi: 10.1016/j.asoc.2014.01.029.

[2] D. J. McFarland and J. R. Wolpaw, "Sensorimotor rhythm-based brain-computer interface (BCI): Feature selection by regression improves performance," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 3, pp. 372–379, Sep. 2005, doi: 10.1109/TNSRE.2005.848627.

[3] X. Wan *et al.*, "A Review on Electroencephalogram Based Brain Computer Interface for Elderly Disabled," *IEEE Access*, vol. 7, no. c, pp. 36380–36387, 2019, doi: 10.1109/ACCESS.2019.2903235.

[4] J. Jin *et al.*, "Bispectrum-Based Channel Selection for Motor Imagery Based Brain-Computer Interfacing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 10, pp. 2153–2163, 2020, doi: 10.1109/TNSRE.2020.3020975.

[5] Z. A. A. Alyasseri *et al.*, "EEG Channel Selection for Person Identification Using Binary Grey Wolf Optimizer," *IEEE Access*, vol. 10, pp. 10500–10513, 2022, doi: 10.1109/ACCESS.2021.3135805.

[6] Pawan and R. Dhiman, "Electroencephalogram channel selection based on pearson correlation coefficient for motor imagery-brain-computer interface," *Meas. Sensors*, p. 100616, Dec. 2022, doi: 10.1016/j.measen.2022.100616.

[7] S. G and S. K. J, "Artificial Neural Network Based Classification Of Motor Imagery EEG Signals For Efficient Brain Computer Interface System," *2023 International Conference on Recent Trends in Electronics and Communication (ICRTEC)*, Mysore, India, 2023, pp. 1-4, doi: 10.1109/ICRTEC56977.2023.10111856.

[8] J. R. Albert *et al.*, "Investigation on load harmonic reduction through solar-power utilization in intermittent SSFI using particle swarm, genetic, and modified firefly optimization algorithms," *J. Intell. Fuzzy Syst.*, vol. 42, no. 4, pp. 4117–4133, 2022, doi: 10.3233/JIFS-212559.

[9] S. Aggarwal and N. Chugh, "Signal processing techniques for motor imagery brain computer interface: A review," *Array*, vol. 1–2, no. June, p. 100003, 2019, doi: 10.1016/j.array.2019.100003.

[10] Pawan and R. Dhiman, "Motor imagery signal classification using Wavelet packet decomposition and modified binary grey wolf optimization," *Meas. Sensors*, p. 100553, Dec. 2022, doi: 10.1016/j.measen.2022.100553.

[11] Blankertz, B., Dornhege, G., Krauledat, M., Müller, K. and Curio, G. "The non-invasive Berlin Brain – Computer Interface: Fast acquisition of

Copyright © IEEE–2024 ISBN: 979-8-3503-0644-6

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:46 UTC from IEEE Xplore. Restrictions apply.

effective performance in untrained subjects" *Neuroimage*, 2007, Vol. 37, pp. 539–550

[12] Pawan and R. Dhiman, "Machine learning techniques for electroencephalogram based brain-computer interface: A systematic literature review," *Meas. Sensors*, vol. 28, p. 100823, Aug. 2023, doi: 10.1016/j.measen.2023.100823.

[13] N. Burioka *et al.*, "Approximate Entropy in the Electroencephalogram During Wake and Sleep", *Clin EEG Neurosci*., 2005 Jan, doi: 10.1177/155005940503600106.

[14] A. Hamad, E. H. Houssein, A. E. Hassanien, and A. A. Fahmy, "A hybrid EEG signals classification approach based on grey wolf optimizer enhanced SVMS for epileptic detection," *Adv. Intell. Syst. Comput.*, vol. 639, no. January, pp. 108–117, 2018, doi: 10.1007/978-3-319-64861-3_10.

[15] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014, doi: 10.1016/j.advengsoft.2013.12.007.

[16] J. Too, A. R. Abdullah, N. M. Saad, N. M. Ali, and W. Tee, "A new competitive binary grey wolf optimizer to solve the feature selection problem in EMG signals classification," *Computers*, vol. 7, no. 4, 2018, doi: 10.3390/computers7040058.

[17] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, 2016, doi: 10.1016/j.neucom.2015.06.083.

[18] S. Huang, C. A. I. Nianguang, P. Penzuti Pacheco, S. Narandes, Y. Wang, and X. U. Wayne, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics and Proteomics*, vol. 15, no. 1. International Institute of Anticancer Research, pp. 41–51, Jan. 01, 2018. doi: 10.21873/cgp.20063.

[19] Y. Tang and D. M. Durand, "A tunable support vector machine assembly classifier for epileptic seizure detection," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 3925–3938, 2012, doi: 10.1016/j.eswa.2011.08.088.

[20] I. Güler and E. D. Übeyli, "Multiclass support vector machines for EEG-signals classification," *IEEE Trans. Inf. Technol. Biomed.*, vol. 11, no. 2, pp. 117–126, Mar. 2007, doi: 10.1109/TITB.2006.879600.

Copyright © IEEE–2024 ISBN: 979-8-3503-0644-6

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 19,2024 at 05:54:46 UTC from IEEE Xplore. Restrictions apply.

2024 5th International Conference for Emerging Technology (INCET)
Karnataka, India. May 24-26, 2024

Multifaceted Discrete Emotion Recognition from EEG Physiological Signals via Machine Learning Techniques

Amriti Gupta
Dept. of Software Engineering
Delhi Technological University
amritigupta_mc21a11_19@dtu.ac.in

Kunal Nayyar
Dept. of Electronics & Communication Engineering
Delhi Technological University
kunalnayyar_ec21a17_19@dtu.ac.in

Harshit Chopra
Dept. of Computer Engineering
Delhi Technological University
harshitchopra_co21a4_13@dtu.ac.in

Neha Gahlan
Dept. of Software Engineering
Delhi Technological University
nehagahlan_2k21phdcs02@dtu.ac.in

Divyashikha Sethia
Dept. of Software Engineering
Delhi Technological University
divyashikha@dtu.ac.in

2024 5th International Conference for Emerging Technology (INCET) | 979-8-3503-6115-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/INCET61516.2024.10592936

***Abstract*—Emotions, intricate mental states shaped by neurophysiological alterations influenced by cognitive processes, sensory inputs, behaviours, and diverse experiences, remain a compelling domain for exploration. The Electroencephalogram (EEG) is a potent tool that directly quantifies these emotional variations by capturing brain signals. Recent strides in emotion recognition have harnessed traditional machine learning classifiers to automate the identification of human emotions with remarkable success. This paper delves into the relatively underexplored Discrete Emotion Model, unveiling its capacity to achieve outstanding accuracy despite historical reservations regarding its effectiveness, using the ECSMP (Emotion, Cognition, Sleep, and Multi-model Physiological signals) dataset. It has two distinct environments, Video watching, and CANTAB-based cognitive assessment phases, enabling seamless data collection and analysis. This research effectively quantifies emotion for binary classification (classifies the type of emotion felt) and multiclass classification (classifies the intensity of emotion felt), elevating emotion recognition capabilities through the synergy of EEG technology. The exceptional performance of XGBoost, with a 96.5% accuracy rate in binary classification and 95% in multiclass emotion recognition, highlights its prowess compared to the other models tested.**

***Index Terms*—Emotion Recognition, Physiological signals**

I. Introduction

Emotions are states experienced by individuals that influence their behaviours and thoughts. They arise from both physical and physiological reactions to internal or external stimuli. Differentiating between emotions can be done through facial expressions, behaviours, and physiological responses. Physiological signals are precious in accurately reflecting a person's emotions in real-time. Monitoring these signals has become more accessible through wearable devices such as smartphones and smartwatches. However, improved emotional well-being necessitates developing techniques to extract emotional states from physiological measurements acquired via mobile devices. Electroencephalogram (EEG) is a widely employed and effective physiological signal for discerning emotions [1], [2]. This efficacy arises from its direct capture of signals emanating from the brain.

A. Emotion Model

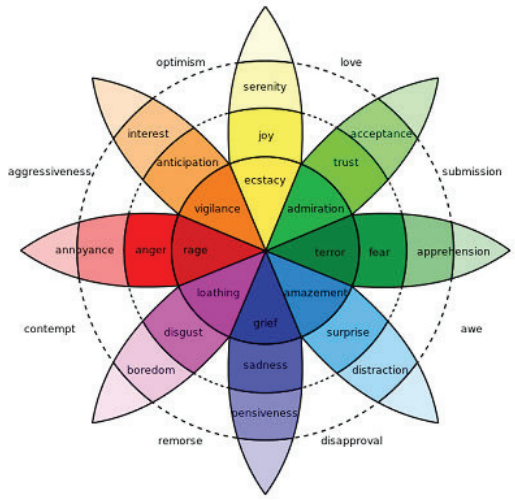


Fig. 1: Plutchik's Emotion Model [3]

Scholars and psychologists have proposed two prominent and widely accepted theories in studying emotions: the Discrete Emotion Model and the Dimensional Emotion Model [4], [5].

From a discrete perspective, influenced by Ekman's theory [15] and evolutionary insights from Darwin and Tomkins, six primary emotions emerge: happiness, sadness, anger, disgust, fear, and surprise. Additional feelings result from intricate combinations or variations of these core emotions. Plutchik [16] introduces a wheel model as shown in Fig. 1, classifying eight primary emotions by intensity. It is similar to mixing primary colors to create secondary hues. Izard [17]

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

TABLE I: Related work for Emotion Recognition using Discrete Emotion Model using EEG

Author	Emotion Labels	Subjects	Feature Extraction	Classifiers	Accuracy (in %)
Murugappan et al [6]	5 emotions (anger, disgust, happy, surprise, sad)	5	Wavelet based features	MLP	joy: 52.04, anger: 50, relax: 33.4, sad: 25, fear: 50.1
Ranganathan et al [7]	7 emotions (angry, happy, sad, disgust, fear, surprise, neutral)	10	Spectral features	DBN	79.2 (average)
Takahashi [8]	5 emotions (joy, anger, sad, fear, relaxation)	12	Time Domain Features	Linear SVM, NN	joy: 52, anger: 50, relax: 33.4, sad: 25, fear: 50.1
Petrantonakis et al [9]	6 emotions (happy, surprise, anger, fear, disgust, sad)	16	Statistical features, WT, HOC based features	QDA, KNN, SVM	83.33 (average)
Nasehi et al [10]	6 emotions (happy, surprise, anger, fear, disgust, sad)	10	Gabor-PCA based features	PNN Classifier	happy: 58.76, surprise: 67.05, anger: 73.64, fear: 56.79, disgust: 69.47, sad: 62.97
Khosrowabadi et al [11]	4 emotions (calm, happy, sad, fear)	26	Time and Frequency Domain features	KNN, SVM	58.26 (average)
Bhatti et al [12]	4 emotions (happy, love, sad, anger)	30	Time, Frequency and Wavelet Domain Features	MLP, KNN, SVM	happy: 94.87, love: 65.38, sad: 78.125, anger: 74.07
Liu et al [13]	8 emotions (joy, amusement, anger, tenderness, sadness, fear, disgust, neutrality)	30	Time-Frequency Domain Features	SVM, LIBSVM	60.55 (average)
Murugappan et al [14]	5 emotions (happy, disgust, fear, surprise, neutral)	25	Absolute Logarithmic Recoursing Energy Efficiency	KNN, LDA	happy: 81.67, surprise: 81.67, fear: 81.25, disgust: 91.67, neutral: 93.75

identifies ten fundamental emotions, and Parrott [18] offers a comprehensive classification encompassing various emotional states.

In contrast, the dimensional model divides emotional space into two dimensions: Valence-Arousal (VA) or three dimensions with the Pleasure-Arousal-Dominance (PAD) framework. Valence encompasses pleasure or displeasure, spanning positive and negative aspects. Arousal signifies emotional alertness and physiological intensity, while Dominance denotes control or external influence.

The dimensional model assigns numerical values to represent emotions across the valence-arousal spectrum, even in complex cases. However, it faces challenges when similar feelings cluster closely on the valence-arousal plane, yielding distinct representations and behavioural reactions [19]. Researchers often categorize valence and arousal attributes in emotion recognition rather than pinpointing specific emotions. Emotions within the same category can vary in intensity, as with happiness from a sunset or academic success. Thus, while joy can have diverse valence and arousal levels, the reverse is not necessarily true.

B. Motivation

This work is motivated by the limited prior research in discrete model emotion recognition, noting that existing studies have not consistently demonstrated high precision, despite the potential superiority of the discrete model in handling complex emotions compared to the dimensional model. Addressing this knowledge gap and the opportunity to improve complex emotion recognition using the discrete model have been key drivers behind this research.

Table I illustrates the prevalent use of relatively small subject groups in previous studies. In contrast, this study is motivated to adopt a distinctive approach involving more subjects for a more comprehensive and in-depth analysis. The Dataset on Emotion, Cognition, Sleep, and Multi-model Physiological Signals (ECSMP) [20] serves as the basis for this exploration and has received limited prior attention.

Furthermore, the motivation also arises from the under-explored area of emotional intensity identification, often overlooked in previous research. This work aims to develop a comprehensive approach considering emotion recognition and nuanced emotional intensity measurement.

This study also aims to expand research by exploring various subjects and scenarios to assess the generalizability of findings and identify environments that favor effective emotion recognition. Context and conditions significantly impact recognition efficacy, further motivating this exploration. In particular, this study seeks to investigate two distinct environments within the ECSMP dataset [20]: 'CANTAB-Video' and 'Video-CANTAB.' By analyzing these environments, the goal is to determine which one demonstrates superior performance and applicability for emotion recognition tasks. This comparative analysis will contribute valuable insights into the suitability of different environments for such tasks, offering practical implications for future research and applications.

C. Related Work

Numerous studies have delved into the Valence-Arousal-Dominance (VAD) model of emotions, which has garnered significant attention in emotion recognition. However, there needs to be more research focusing on the Discrete Emotion Model. In Table I, a comprehensive summary of related works explicitly about the Discrete Emotion Model is provided, highlighting the limited existing research in this area. Notably, the best average accuracy achieved in previous works stands at 86% using KNN, as demonstrated by Muruguppan et al. [14], signifying room for improvement. This paper aims to advance the less-explored Discrete Emotion Model, addressing a research gap and boosting emotion recognition and classification accuracy.

In the domain of emotional classification, previous research primarily relied on stimuli to categorize participants' emotional states. However, this study illuminates the nuanced nature of human emotions, acknowledging that individual differences, cultural factors, and overall well-being can cause

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

emotional responses to vary, even when participants encounter identical stimuli. Babushkin et al. [21] explored this intricate variation in emotional intensity, achieving a 70% accuracy rate.

D. Contributions

- The study utilizes various ML models like XGBoost, ANN, LSTM, CNN, and KNN to fit the discrete emotion recognition model, fine-tuning them results in XGBoost outperforming every other model.
- Two types of environments were analyzed, namely, Cantab-Video and Video-Cantab. Cantab-Video gave significantly better results after training, underlining the importance of ambience, surroundings and temperament.

These findings advance the state-of-the-art in emotion recognition and open doors to more nuanced emotional intensity recognition, offering valuable contributions to fields such as psychology, healthcare, and human-computer interaction.

E. Roadmap

The paper comprises several sections. Section II presents the methodology of the proposed work, Section III focuses on the evaluation and discussion of results, and Section IV concludes the paper, providing insights into future research directions.

II. Methodology

The flowchart for the methodology being used in this study is shown in Fig. 2.

A. Dataset Description

The ECSMP [20] dataset comprises EEG data from 73 healthy individuals (26 males, 47 females; mean age: 23.53 ± 2.06 years) recorded during sleep, except for 7 participants who were registered post-experiment due to personal reasons.

During the experiment, participants completed cognitive tasks and watched emotion-inducing videos in a dim, quiet room. The order of the two studies was random. Before conducting both studies, the authors recorded a three-minute resting state with closed eyes followed by three minutes with open eyes to establish a baseline for unbiased recordings.

In the video-watching task, participants watched six 5-minute clips inducing emotions (neutral, fear, sad, happy, anger, disgust) with a 120-second break between each video. Post-viewing, the participants were required to fill out the Current Emotional Self-Assessment Form (CESAF) to describe their current emotion intensity level on a scale of 1-7 after watching videos for 30 seconds to assess arousal of each emotion, followed by a 90-second rest with eyes closed to recover from the previously induced feeling.

Participants' prefrontal and ear electrodes recorded scalp EEG. The dataset contains EEG records from 87 healthy students (31 males, 56 females; mean age: 23.59 ± 2.02 years) and responses from 88 subjects (31 males, 57 females; mean age: 23.69 ± 2.14 years) on Emotion Regulation Questionnaire (ERQ) , Self-rating Depression Scale (SDS) , Profile Of Mood States questionnaire (POMS) [6], Pittsburgh Sleep Quality Index (PSQI) scales.

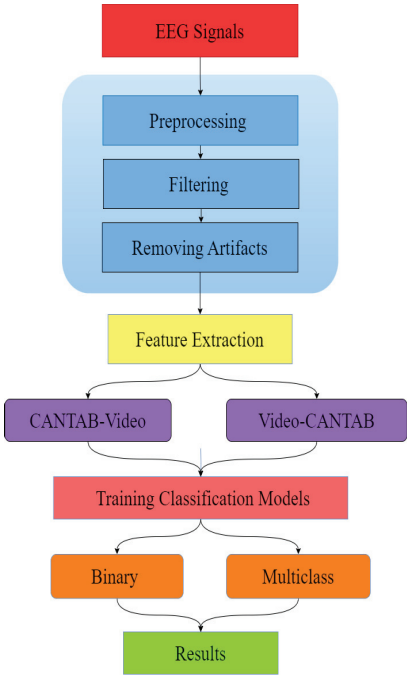


Fig. 2: Block Diagram for Proposed Methodology

B. Experimental Scenarios

Two different scenarios serve as the testing ground for the models. Users fall into two subsets, helping to delve into the nature and extent of emotional expression in humans:

1) In the first subset, the CANTAB experiment preceding the Video Watching task with 27 users.
2) In the second subset, the Video Watching task preceding the CANTAB Experiment, involving 42 users.

The difference between both subsets of user signals is because, several CANTAB-Video users are having a significant amount of noise in the signals after the pre-processing and signal cleaning phase. Hence, we removed it to maintain the integrity and quality of results of this study.

C. Pre Processing

The proposed study adopts the following steps to remove artifacts from the EEG signals, clip the data, and decompose the signal into IMFs for further classification.

1) **Artefact Removal:** Butterworth Bandpass filters achieve a smooth frequency response in the desired frequency range. It is a filter with an optimal magnitude response. This filtering approach ensured a consistent frequency response within the desired range (passband) while gradually reducing responses in the unfrequented range (known as stopband). The pace of this response varies directly with the filter's order. A filter of second order has a -12 dB decrease per octave, third order -18 dB, and so on. Our application implements a second-order filter with a low-cut frequency set at 1Hz and a high-cut frequency set at 40Hz.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

2) **Clipping:** The ECSMP [20] dataset includes EEG signal data from 89 users and demographic details like age, gender, Sleep Stats, and User ID. Initially stored in ".mat" format, the data is downsampled from 1000Hz to 250Hz and converted to ".csv." There are 250 rows per second for each user during the experiment.
The EEG data covers two tasks: cognitive assessment and video-watching. Since the order of these tasks is random for users, EEG signals recorded during cognitive assessments are excluded due to varying completion times, leading to different row counts in the CSV file. This research paper aims to develop models adept at figuring out the emotions of the users based on a "Video Stimulus", CANTAB shifts the user's active brain regions from Amygdala to the prefrontal Cortex, one handles emotions and the other cognitive ability and reasoning, Hence it swerves the user's brain as far as the scope of this paper goes, Although to include the CANTAB EEG Signals and analyze the changes from start to end could be a topic for another paper itself. The EEG device employs seven channels on the forehead: FP1, FZ, FP2, AF3, AF4, F7, and F8.
Preceding the experiment, a six-minute baseline EEG signal is recorded, constituting 90,000 rows. The baselines are extracted for each user and subtracted from the post-stimulus period of EEG Signals to ensure normalization throughout the user database. The investigation entails six videos, each lasting 5 minutes (75,000 rows), followed by participants filling out the CESAF form for 30 seconds and taking 90 seconds to transition to a neutral state before the following video. This cycle includes 120 seconds of rest (30,000 rows) between videos, requiring subsequent clipping. Therefore, this study removes the baseline EEG signals of 90,000 rows and the EEG signals comprising the rest between the videos (30000*6) rows.
3) **Empirical Mode Decomposition (EMD):** EMD, or Empirical Mode Decomposition [22], is an adaptable technique that dissects data and signals into meaningful components. Ideal for non-linear and non-stationary signals, EMD separates them into varying resolution components. It succinctly decomposes complex datasets into a comprehensive set of intrinsic mode functions (IMFs), offering an orthogonal basis for the original signal. The primary IMF usually captures high-frequency elements, enabling the removal of components like random noise. The recursive process of EMD extracts different resolutions directly from the data itself without resorting to fixed functions or filters. Empirical Mode Decomposition (EMD) extracts the clipped EEG signal's first intrinsic mode function (IMF), isolating the most prominent and relevant oscillatory patterns for further analysis and feature extraction.

D. Feature Extraction

This study extracts nine features from pre-processed EEG data across the seven channels, resulting in sixty-three features.

Time Domain Features: Consider the input signal as $x(t)$. This study calculates the time domain features as follows:

1) **Mean (Average):** The mean of the signal $x(t)$ can be computed as:
$$\mu = \frac{1}{N}\sum_{i=1}^{N} x(i) \tag{1}$$
2) **Standard Deviation:** The standard deviation of the signal $x(t)$ measures the extent of values around the mean.
$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x(i)-\mu)^2} \tag{2}$$
3) **Variance:** Variance is the square of the standard deviation, representing the average of squared deviations from the mean.
$$\text{Variance} = \frac{1}{N}\sum_{i=1}^{N}(x(i)-\mu)^2 \tag{3}$$
4) **Hjorth Mobility:** Hjorth's mobility assesses waveform dynamics by comparing variances between the first derivative and the signal.
$$\text{Mobility} = \frac{\sqrt{\text{Variance of the first derivative}}}{\sqrt{\text{Variance of the signal}}} \tag{4}$$
5) **Hjorth Complexity:** Hjorth's complexity measures waveform dynamics by comparing variances in the first and second derivatives.
$$\text{Complexity} = \frac{\text{Variance of the second derivative}}{\text{Variance of the first derivative}} \tag{5}$$

Frequency Domain Features: In the spectral domain, represented by X, these characteristics offer valuable insights into signal intricacy and irregularity through entropy-based measures:

1) **Spectral Entropy:** Measures the uncertainty in power distribution across frequency components.
$$\text{Spectral Entropy} = -\sum_{i=1}^{N} p_i \log_2 p_i \tag{6}$$
where p_i is the normalized power of the i-th frequency component.
2) **SVD Entropy:** Quantifies information in singular values obtained from signal spectrogram matrix.
$$\text{SVD Entropy} = -\sum_{i=1}^{N} \sigma_i^2 \log_2 \sigma_i^2 \tag{7}$$
where σ_i is the i-th singular value.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

TABLE II: BINARY CLASSIFICATION (Accuracy in %)

(a) CANTAB-VIDEO

Emotions	Neutral	Fear	Sad	Happy	Anger	Disgust
XGBoost	96.57	97.77	95.87	96.44	96.90	95.45
CNN	96.18	95.56	96.20	90.06	95.66	88.48
KNN	88.71	80.06	77.92	72.50	81.41	67.81
ANN	95.04	95.00	93.73	91.68	94.93	85.73
LSTM	94.72	95.31	93.51	89.33	94.45	82.47

(b) VIDEO-CANTAB

Emotions	Neutral	Fear	Sad	Happy	Anger	Disgust
XGBoost	87.25	84.57	91.39	90.99	90.84	91.43
CNN	95.67	91.10	91.40	93.21	92.59	93.22
KNN	73.33	65.15	68.86	63.63	66.16	73.33
ANN	84.25	74.73	91.99	90.13	88.52	91.64
LSTM	83.06	73.33	92.52	90.49	87.08	93.22

TABLE III: MULTICLASS CLASSIFICATION (Accuracy in %)

(a) CANTAB-VIDEO

Emotions	Neutral	Fear	Sad	Happy	Anger	Disgust
XGBoost	94.62	94.45	95.48	95.85	95.43	95.46
CNN	87.25	89.92	92.87	85.03	92.63	74.76
KNN	78.29	62.43	64.88	58.98	67.35	57.57

(b) VIDEO-CANTAB

Emotions	Neutral	Fear	Sad	Happy	Anger	Disgust
XGBoost	83.88	89.90	88.94	88.53	88.05	89.41
CNN	65.48	59.37	87.66	38.31	76.68	87.32
KNN	66.95	46.22	64.88	40.27	47.35	59.63

3) **Petrosian Fractal Dimension:** Evaluates signal self-similarity, considering its fractal intricacy.

$$\text{Petrosian FD} = \frac{\log(N)}{\log(N) + \log(N/\delta N)} \tag{8}$$

where N is signal length, and δN is sign change count in the first derivative.

4) **Katz Entropy:** Captures fluctuations' unevenness in signals, incorporating local and global aspects.

$$\text{Katz Entropy} = \frac{\log(L)}{\log(d/L) + \log(N)} \tag{9}$$

where L is the total path length, d is Euclidean distance between start and end points, and N represents signal length.

TABLE IV: Parameters used in various classifiers

Classifiers	Parameters
XGBoost	Booster: gbtree, Learning Rate: 0.3, Max Depth: 6, Min Child Weight: 1, Num Estimators: 100
CNN	Architecture: Conv1D (128, kernel=3, relu), MaxPooling1D (pool=2), Conv1D (128, kernel=3, relu), MaxPooling1D (pool=2), Conv1D (64, kernel=3, relu), MaxPooling1D (pool=2), Dense (64, tanh), Dropout (0.2), Dense (32, tanh), Dropout (0.2), Dense (16, relu), Dropout (0.2), Dense (num_classes, softmax) Training: Loss='sparse_categorical_crossentropy', Adam, Batch=256, Epochs=200, Early Stopping (0.005, patience=5, restore_best_weights=True), Validation=20%
KNN	Number of Neighbours: 11, Distance Metric: Euclidean, Nearest Neighbour Search Method: kdtree
ANN	Architecture: Dense (256, relu), Dense (128, relu), Dropout (0.25), Dense (64, relu), Dropout (0.5), Dense (32, relu), Dense (1, sigmoid) Training: Loss='binary_crossentropy', Adam, Batch=32, Epochs=5, Early Stopping (0.01, patience=5, restore_best_weights=True), Validation=20%
LSTM	Hidden Units: 10, Optimizer: Adam, Epochs: 10, State Activation: tanh, Gate Activation: sigmoid, Input Weights Initializer: glorot, Recurrent Weights Initializer: orthogonal, Bias Initializer: unit-forget-gate

E. Classification Method

This study performs two types of classification, as follows:

1) **Multiclass classification** aims to quantify the intensity of emotions on a scale of 1 to 7, with 1 being the lowest and 7 being the highest. The model learns to predict the power of the emotion based on the EEG signal. As we train the model on each emotion, the model only needs to learn about the trends of a specific feeling at a time. The loss function minimizes the error between the ratings provided by the users and the intensity predicted by the model.
2) **Binary Classification** is derived from MultiClass Classification employing putting a threshold value below which the output indicates a weak emotion and above which it shows a strong emotion. The proposed study uses a threshold value of 4 on a scale of 1 to 7 due to being in the middle of the intensity scale.

F. Classification Algorithms

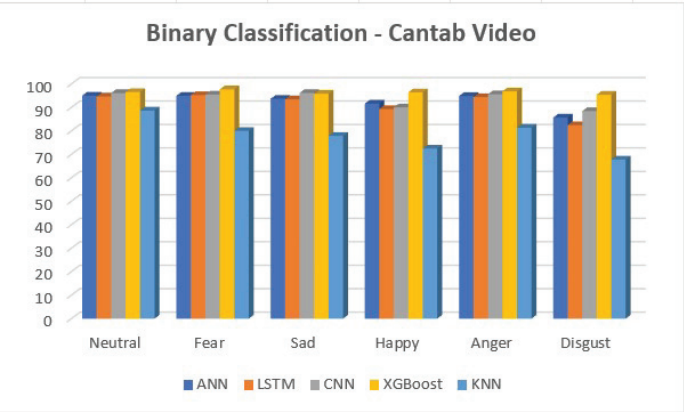
After the feature extraction process, the study uses the input and output files to train the classification models. It distinguishes between 'CANTAB-Video' and 'Video-CANTAB' by testing various classifiers to suit the dataset.

For binary classification, the study uses Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, XGBoost, and Convolutional Neural Network (CNN) models on the pre-processed dataset featuring EEG signal-derived features. For the multiclass classification, the study uses XGBoost, CNN, AdaBoost, and k-Nearest Neighbors (KNN) classifiers to predict the emotional intensity levels on a scale of 1 to 7.

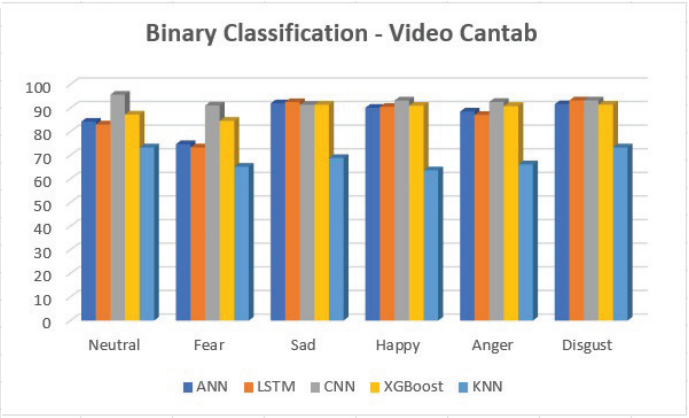
The paper employs the following classification algorithms, along with their corresponding parameters described in Table IV, to train the dataset as elucidated below:

*1) **XGBoost**:* Its basis is the Gradient Boosting framework, an ensemble of decision trees that combine to derive complex relationships between data. This model includes L1 (Lasso) and L2 (Ridge) Normalization to prevent overfitting and bias of the data and enhance the overall.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

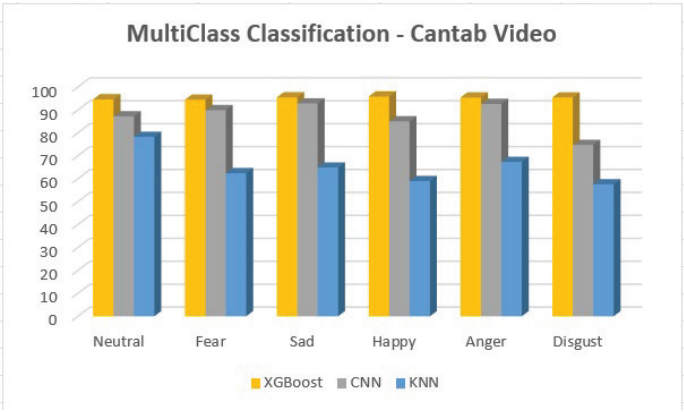


(a) Results for CANTAB-Video Binary Classification

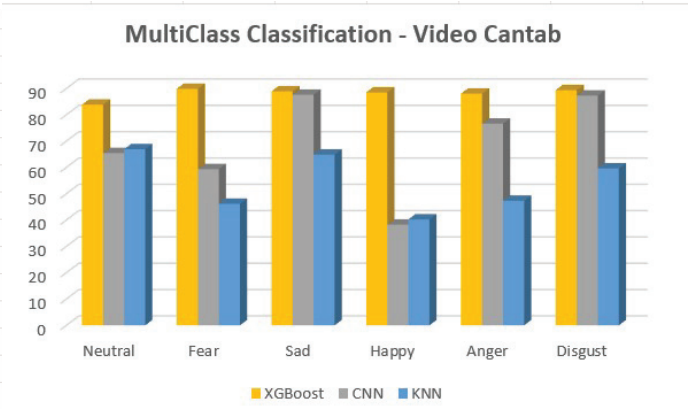


(b) Results for Video-CANTAB Binary Classification

Fig. 3: Binary Classification Results



(a) Results for CANTAB-Video MultiClass Classification



(b) Results for Video-CANTAB MultiClass Classification

Fig. 4: Multiclass Classification Results

*2) **AdaBoost**:* In multiclass EEG signal classification, 500 models (weak learners) are trained with a learning rate 0.1 using the AdaBoost ensemble learning boosting algorithm. The final step combines these models by calculating a weighted average of their predictions.

*3) **Convolutional Neural Networks (CNN)**:* 1D ConvNets extract relevant features from one-dimensional input using kernel-based convolution. We have implemented our custom CNN model, which has 106K trainable parameters.

*4) **K-Nearest Neighbours (KNN)**:* K-Nearest Neighbour (KNN) is a supervised machine learning algorithm using a lazy learning approach. It classifies new data points based on Euclidean distance from the training set data points, selecting the nearest neighbor (K=1) in multiclass EEG signal classification.

*5) **Artificial Neural Networks (ANN)**:* This architecture comprises interconnected artificial neurons and consists of three layers: Input, Hidden, and Output. Each layer processes the signal and passes it to the next layer. This work uses techniques like Dropout and Batch Normalization to foster neuron independence and enhance the model.

*6) **Long short-term memory (LSTM) network**:* LSTMs address RNNs' challenges with long-range dependencies. An LSTM consists of LSTM units with three gates: the input gate, the forget gate, and the output gate. The forget gate's crucial role is determining which information to retain before progressing to the next LSTM unit. Training involves Stochastic Gradient Descent and updates through backpropagation over time.

III. Results and Discussion

In the context of binary emotion classification tasks within the CANTAB-Video setting, the study evaluates five models—ANN, LSTM, CNN, KNN, and XGBoost—for their performance. These models exhibit remarkable accuracies, as depicted in both Fig. 3 and Table V. Notably, XGBoost surpasses all others, achieving an outstanding accuracy of 96.5%.

In transitioning to the Video-CANTAB scenario, the study employs the same five models for binary emotion classification tasks. Notably, there is a discernible difference in the results regarding average accuracies, as evident from Fig. 3 and Table

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

TABLE V: Binary Classification Results (Accuracy in %)

Environment	ANN	LSTM	CNN	KNN	XGBoost
CANTAB-Video	94.3	94.6	94.5	78.1	96.5
Video-CANTAB	85.0	84.7	90.5	68.4	88.5

TABLE VI: Multiclass Classification Results (Accuracy in %)

Environment	CNN	KNN	XGBoost
CANTAB-Video	87.0	64.9	95.0
Video-CANTAB	71.0	54.2	88.0

V. XGBoost stands out, achieving an accuracy of 88.5%, showcasing its superior performance compared to the other models.

In the context of multiclass emotion classification within the CANTAB-Video setting, the study assesses three models—XGBoost, CNN, and KNN—for their performance. These models achieve the following average accuracies, as shown in Fig.4 and Table VI. XGBoost achieves an impressive accuracy of 95.0%. CNN exhibits lower average accuracies, particularly for emotions like happiness and fear, with an accuracy of 87%. KNN achieves an accuracy of 64.9%.

As we transition to the Video-CANTAB scenario, we evaluate the same three models for multiclass classification. All models show a collective decrease in average multiclass accuracy compared to CANTAB-Video, as evident from Fig. 4 and TableVI. XGBoost outperforms the other models with an accuracy of 88%. CNN follows with an accuracy of 71%, while KNN achieves an accuracy of 54.2%. Once again, we highlight the impact of task sequencing on emotion classification performance through the collective decrease in average accuracies compared to CANTAB-Video.

IV. Conclusion

Emotions can manifest in individuals of all backgrounds, leading to variations in behaviour and physiological responses that may influence overall well-being. This study explores the Discrete Emotion Model's intricacies, achieving remarkable accuracy despite historical observations of lower rates. This paper also significantly contributes by accurately measuring the intensity of emotions using the multiclass scenario, thus advancing emotion recognition capabilities. XGBoost excels in this study's experiments, achieving 96.5% accuracy in binary and 95% in multiclass emotion recognition. By investigating two distinct environments, "CANTAB-Video" and "Video-CANTAB," this research reveals that the "CANTAB-Video" environment consistently offers superior outcomes for emotion recognition. These results underscore the importance of environment selection in refining recognition accuracy.

Future research in this domain should explore different feature combinations to enhance the system's classification accuracy. Moreover, this research can extend to exploring intricate interactions among emotion, cognition, and sleep using the ECSMP [20] dataset, resulting in a more comprehensive understanding of their impact on EEG signals and valuable insights for emotion recognition applications.

References

[1] N. Gahlan and D. Sethia, "Federated learning inspired privacy sensitive emotion recognition based on multi-modal physiological sensors," *Cluster Computing*, pp. 1–23, 2023.

[2] ——, "AFLEMP: Attention-based Federated Learning for Emotion recognition using Multi-modal Physiological data," *Biomedical Signal Processing and Control*, vol. 94, p. 106353, 2024.

[3] D. Chafale and A. Pimpalkar, "Review on developing corpora for sentiment analysis using plutchik's wheel of emotions with fuzzy logic," *International Journal of Computer Sciences and Engineering (IJCSE)*, vol. 2, pp. 14–18, 01 2014.

[4] N. Gahlan and D. Sethia, "Three dimensional emotion state classification based on EEG via empirical mode decomposition," in *2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1)*. IEEE, 2023, pp. 1–6.

[5] M. A. Anwar, M. Agrawal, N. Gahlan, D. Sethia, G. K. Singh, and R. Chaurasia, "FedEmo: A privacy-preserving framework for emotion recognition using EEG physiological data," in *2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2023, pp. 119–124.

[6] P. Murugappan, M. Juhari, R. Nagarajan, and S. Yaacob, "An investigation on visual and audiovisual stimulus based emotion recognition using EEG," *International Journal of Medical Engineering and Informatics*, vol. 1, 01 2009.

[7] H. Ranganathan, S. Chakraborty, and S. Panchanathan, "Multimodal emotion recognition using deep learning architectures," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1–9.

[8] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," *2nd International Conference on Autonomous Robots and Agents*, 01 2004.

[9] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 186–197, 2010.

[10] S. Nasehi and H. Pourghassem, "An optimal EEG-based emotion recognition algorithm using gabor features," *WSEAS Transactions on Signal Processing*, vol. 8, pp. 87–99, 01 2012.

[11] R. Khosrowabadi, M. Heijnen, A. Wahab, and H. C. Quek, "The dynamic emotion recognition system based on functional connectivity of brain regions," in *2010 IEEE Intelligent Vehicles Symposium*, 2010, pp. 377–381.

[12] A. M. Bhatti, M. Majid, S. M. Anwar, and B. Khan, "Human emotion recognition and analysis in response to audio music using brain signals," *Computers in Human Behavior*, vol. 65, pp. 267–275, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0747563216305945

[13] Y.-J. Liu, M. Yu, G. Zhao, J. Song, Y. Ge, and Y. Shi, "Real-time movie-induced discrete emotion recognition from EEG signals," *IEEE Transactions on Affective Computing*, vol. PP, pp. 1–1, 01 2017.

[14] P. Murugappan, N. Ramachandran, and Y. Sazali, "Classification of human emotion from EEG using discrete wavelet transform," *J. Biomedical Science and Engineering*, vol. 334054, pp. 390–396, 01 2010.

[15] P. Ekman, "An argument for basic emotions," *Cognition and Emotion*, vol. 6, no. 3-4, pp. 169–200, 1992. [Online]. Available: https://doi.org/10.1080/02699939208411068

[16] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," *American Scientist*, vol. 89, no. 4, pp. 344–350, 2001. [Online]. Available: http://www.jstor.org/stable/27857503

[17] C. Izard, "Basic emotions, natural kinds, emotion schemas, and a new paradigm," *Perspectives on psychological science*, vol. 2, p. 260, 09 2007.

[18] W. G. Parrott, "Emotions in social psychology : essential readings," 2001. [Online]. Available: https://api.semanticscholar.org/CorpusID:141721437

[19] D. Yin, S. Bond, and H. Zhang, "Anxious or Angry? effects of discrete emotions on the perceived helpfulness of online reviews," *MIS Quarterly*, 05 2013.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

[20] Z. Gao, X. Cui, W. Wan, W. Zheng, and Z. Gu, "ECSMP: A dataset on emotion, cognition, sleep, and multi-model physiological signals," *Data in Brief*, vol. 39, p. 107660, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340921009355

[21] V. Babushkin, W. Park, M. Jamil, H. Alsuradi, and M. Eid, "EEG-based classification of the intensity of emotional responses," 05 2021, pp. 218–221.

[22] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998. [Online]. Available: http://www.jstor.org/stable/53161

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:29:10 UTC from IEEE Xplore. Restrictions apply.

Journal of Material Cycles and Waste Management
https://doi.org/10.1007/s10163-024-02018-1

ORIGINAL ARTICLE

Municipal organic solid waste management in the concept of urban mining and circular economy: a model from Palestine

Majed Ibrahim Al-Sari'[1,2] · A. K. Haritash[1]

Received: 19 May 2022 / Accepted: 25 June 2024
© The Author(s), under exclusive licence to Springer Nature Japan KK, part of Springer Nature 2024

Abstract

Municipal organic solid waste in Palestine is causing several environmental and social problems. This paper studied management of this waste stream in the concept of urban mining and circular economy through composting and reuse the produced compost for different purposes as an indicator for the circular economy through the period extended from 2021 up to 2035. Two compost reuse scenarios are studied: use for agricultural purposes, and use as landfill cover materials. For agricultural purposes, compost can be used as an alternative fertilizer where nutrients in compost can replace the existing chemical fertilizers available in Palestine such as humic acid as a source of carbon, Ammonium Sulphate as a source of nitrogen, Triple Super Phosphate as a source of Phosphorus, and Potassium Phosphate source of Potassium. The estimated revenue from compost use in agriculture is USD 194.8 million in 2021 and USD 369.8 million in 2035. The estimated saving from using compost as landfill cover is USD 0.876 million annually, and USD 13.14 million during the 15 years' study period. Implementation of the circular economy principles in municipal solid waste management through composting can close the materials recycling loop, generating extra income, and adding net revenue to the national economy.

Keywords Organic Waste · Compost · Circular Economy · Fertilizers · Municipal Waste

Abbreviations

EU	European Union
ECN	European Compost Network
SWM	Solid Waste Management
C&D	Construction and Demolition
WEEE	Waste from Electric and Electronic Equipment
JSC-H&B	Join Service Council for Solid Waste Management for Hebron and Bethlehem Governorates
USD	United States Dollar
CDM	Clean Development Mechanism
LandGEM	Landfill Gas Emission Model
IFC	International Finance Corporation
NIS	New Israeli Shekel
SRTAC	Sharabati Brothers Trading and Agricultural Company
C	Carbon
N	Nitrogen
K	Potassium
P	Phosphorus
TSP	Triple Super Phosphate
SOP	Potassium Sulphate
OM	Operation and Maintenance
AS	Ammonium Sulphate
KSA	Kingdom of Saudi Arabia
DAP	Diammonium Phosphate
CO2	Carbon Dioxide

Introduction

Circular economy in waste management

The linear economic model "make-use-dispose" is the dominant global economy. It is considered as a challenge because it can generate waste, degrade the environment, cause price hike due to volatile resources and supply disruption, and cause supply risks for countries that depend

✉ Majed Ibrahim Al-Sari'
majedsari@gmail.com

[1] Department of Environmental Engineering, Delhi Technological University (Formerly Delhi College of Engineering), Bawana Road, Shahbad Daulatpur, Delhi 110042, India

[2] Universal Institute for Applied and Health Research (UIAHR), Nablus, Palestine

Published online: 08 July 2024

on imports. However, the circular economy is an alternative to the linear economy which can solve these challenges and promote environmental practices by returning the discarded products into the production cycle. The linear economy is the dominant consumption model in the developing world, while developed countries, such as the European Union (EU), are applying the principles of the circular economy in policy and action including the waste management sector [1]. In accordance with the EU plan, 65% of the municipal waste shall be recycled as a target [2] and food waste per capita shall be reduced on the retailer-consumer level to half by 2030 [3]. The principles of the circular economy include the preservation and enhancement of natural capital by controlling finite stocks and balancing renewable resource flows; optimization of resource yields by circulating materials at the highest utility at all times in both technical and biological cycles; and fostering system effectiveness by revealing and designing out negative externalities through reduced damage to systems (food, mobility, shelter, education, etc.) and managing externalities (land use, air, water, etc.) [4].

The aim of the circular economy is to use the products as longer as possible by reusing and repairing them to reduce waste production and use of more secondary raw materials in the production cycle [5]. Composting of municipal organic waste is an application of the use of secondary materials in the production cycle instead of depositing them at landfills. Following the model of the circular economy, previous studies highlighted the recycling of organic waste materials into various products. For example, grape skin and seeds along with stems can be recycled and converted into compost [6], and an organic fraction of municipal waste such as household food waste being transferred into compost [7]. The compost is rich in carbon, which in the soil, can resist soil erosion, improve soil water retention, improve soil fertility, and increase biodiversity due to the enumeration enrichment of bacteria, fungi, insects, and worms [8], which can support better soil health and plant growth, ultimately boosting its resilience to cope with harsh drought conditions.

Compost can support the food chain as well as food security by returning nutrients to the soil and enhancing its productivity and crop yield [4, 8]. In accordance with the European Compost Network (ECN), bio-waste through compost can contribute to the circular economy objectives via closing biological material cycles, and reducing the linear economy of landfilling of organic waste; rich in carbon and nutrients thus contributing to long-term soil fertility and C-sequestration through the production of quality soil amendments and organic fertilizers; producing bio-based products which can replace fossil-based products such as mineral fertilizers, peat, and fossil fuels [9].

Compost and urban mining

Composting is a technique for urban mining of organic waste to return it back to the material cycles as a form of circular economy. Through this process, the valuable materials are isolated/ extracted from the waste stream, recycled, and returned to the material cycle. It is widely defined in the literature. It is a process in which products, components of products, elements, and secondary raw materials from used products or elements from urban metabolisms are first discarded, and then recovered and reused as products or resources [10]. It is an effective way to recover resources from secondary mining, landfill mining, and hazardous waste deposits and includes hibernation mining, dissipation mining, and in-use mining [11]. Urban mining is, indirectly, the activity of reclaiming compounds, energy, and elements from products, buildings, and waste generated from urban catabolism [12, 13]. The terminology of urban mining was used for exploitation and extraction of secondary materials from landfills in the beginning, but nowadays it is widely used to include any sort of material recycling [5]. Using the terminology of urban mining that includes any sort of material recycling, extraction of organic waste toward its composting can be considered as compost of urban mining. Tejaswini et al. [14] reported the different types of urban mining (urban mined waste from electrical and electronic equipment -WEEE, urban mined construction and demolition (C&D) waste, urban mined plastic waste and urban mined organic waste) and mentioned that composting is an urban mining of solid waste. Moreover, it has been reported that the term "Urban Mining" is applicable to any type of waste stream [15, 16]. The term 'Urban Mining", "Landfill Mining", "Circular Economy" are contributors to waste management and have strong relationships. Since urban mining is extraction of reusable and recyclable material from the waste stream, and landfill mining is the extraction of such materials from old landfills, then these extracted materials are entered the material cycles through the circular economy concept and contribute to waste management.

The benefits of urban mining were widely reported in the literature. For example, mining of WEEE can reduce 80% of the raw materials required for production of new devices [17], source of job opportunities [18], reduce mining from ores and thus reduce environmental and social impacts [19], reduce the cost of material as identified by Zheng et al. [20] who found that the mean urban mining costs of copper and aluminum are less than the costs of virgin mining, urban mining and recycling of one ton of C&D waste reduce dependence on fossil fuel energy and have the potential for generating environmental benefits such as reduction in carbon emissions [21].

The composting process can be used to recover nutrients from the waste material [22], which can be used in agriculture to replace chemical fertilizers. Ddidba et al. [23] studied the circular economy approach for nutrient recovery from organic waste (fecal sludge, sewage sludge and solid waste) in the context of low- and middle-income countries using composting. Composting food waste is an effective method to valorize this waste stream into nutrient-rich biofertilizers [24]. This can achieve environmental and social benefits through reduction in emissions and nuisance, and reducing the energy and virgin raw materials used to produce chemical fertilizers simultaneously. Further, organic waste can be co-composted with other materials such as soybean crude residue and rice husk can enhance biodegradation [25] and return such materials to the material cycles and circular economy. Moreover, some degrading-resistant organic waste can be biochar amended during composting to enhance microbial activity and organic matter biotransformation into compost [26], thus enhancing the biochar material cycles. Previous studies reported that partial substitution of chemical fertilizers by organic matter can provide the soil with enough nutrient and increase yield [27], selenium-rich compost increase soil fertility and selenium content [28, 29] and replace inorganic selenium fertilizers, replacing 30% chemical fertilizers by organic one increase the quality and yield of cabbage plants [30], and replacement of chemical fertilizers by manure and biological fertilizers can achieve a qualitative and quantitative yield on saffron [31].

Potential for compost production in the study area

Organic fraction represents 46% of the total waste stream in the study area [32], which generates 1,200.0 tons of solid waste daily in average. This showed great potential for compost production from the urban mining of the organic fraction of the fresh waste collected on daily basis. The separation of this waste stream at the source can contribute to the production good quality compost and avoid potential contamination of heavy metals.

Another potential source of compost is the landfill mining, which is a type of urban mining, can result in the extraction of decomposed organic matter from the landfill to be reused. A previous study reported that solid waste buried in landfills for a long time can be excavated, extracted and used as compost for plants or can be reused as biocover for actively operating landfills [33], which can enhance methane oxidation by methanotrophic bacteria [34–36]. In the study area, there are 18 random dumpsites closed more than 9 years ago and contain large quantities of organic waste which can be extracted and used as compost. A study on landfill mining in Indonesia based on the garbage heap age found that the best percentage of compost can be attained after the garbage heap reaches the optimum age of 8–9 years [37]. In addition, the current regional landfill in operation in the study area, which contains more than 2.5 million tons of solid waste, is facing political and social constraints for expansion, which makes landfill mining one of the options to extend the service life of the landfills [33]. Moreover, the agricultural sector in Palestine is facing a severe shortage of chemical fertilizers due to the restrictions imposed by Israel on importing high-quality dual-use fertilizers, which makes the composting of organic waste one of the best options in hand.

However, special safety consideration and measures should be taken when planning to use compost produced from landfill mining in agriculture due to heavy metals contamination. Although some studies reported the suitability of soil-like material extracted from old landfills for soil amendment [38, 39], other studies pointed out high content of heavy metals [40–42] and, accordingly, such soil is not suitable for agricultural purposes. The effects of the heavy metals on the environment and human health owing to their toxicity and bioaccumulation were widely reported in the literature. For example, it is characterized by high toxicity and may be lethal to human [43], carcinogenic such as chromium [44], poisonous which causes intellectual abnormalities in children such as lead [45] and changes in synaptic transmission and neurotransmitter balance such as arsenic [46], and mercury causes kidneys' membranous nephropathy [47]. Therefore, the proper use of compost extracted from landfill mining shall be restricted to non-agricultural applications as recommended by Datta et al. [42] such as cover material for waste in active landfills.

Problems of solid waste management

Landfilling is the main waste disposal method currently in use, which poses environmental and health impacts due to the spread of flies and mosquitoes, odor release and greenhouse gas (GHG) emissions. The national reports showed that the emissions from the SW sector reaches 23% of the total emissions in Palestine [48], which is one of the main contributing sectors to the global warming and climate change as well. In accordance with the national strategies and plans, solid waste landfilling shall be reduced by 30% by 20,230 and 50% by 2040, and methane emissions shall be lowered by 30% by 2030 and 40% by 2040 [49]. Previous studies highlighted several negative impacts of landfills on the environment and human health in the surrounding communities. These were related to odor emissions, impacts on livestock and plants, hygienic impacts on nearby residents, leachate spills which causes pollution, impacts on aesthetic views and traffic, reduction in the value of the properties and nearby land, and health impacts related of the landfill of being a source of insects and rodents [50]. Moreover, landfill construction and expansion faces severe social and political

restrictions. A study conducted by Al-Khatib et al. [51] in the study area showed that more than half of the respondents were against landfill construction at 1 km from their houses. In the study area, in particular, urban mining of recyclables including metals, plastics, and wood is very limited and do not exceed 0.5% of the total waste stream while urban mining of organic waste and composting was initiated at pilot scale and stopped due to financial and technical constraints [49]. Therefore, proper management of the organic fraction of municipal solid waste is a pressing issue due to the large quantities of this waste stream and high environmental and social impacts.

Considering the problems of waste management and national plans mentioned above, it becomes imperative to devise ways that can help reduce the extent of the problem, minimize the environmental impacts, and promote sustainability without much of an economic burden. Utilizing the organic fraction of waste can significantly reduce the load over the waste management system and its composting can lead to the transformation of waste into wealth through a natural process. Apart from it, mining of compost from operational/ closed landfills can help create more space for waste landfilling. Since compost has nutrient value, its use can be promoted in agriculture, thereby completing the nutrient cycling and reducing the economic burden in terms of the impact of chemical synthetic fertilizers. Therefore, this study focuses only on the management of organic solid waste fraction separated at the source following the concept of urban mining and circular economy. Compost from landfill mining is not evaluated and not considered in this study due to the absence of reliable data on waste quantities and quality. The purpose of this study is to evaluate the role of composting in the circular economy through two scenarios: (1) the use of compost to replace chemical fertilizers in agriculture and (2) the use of compost as cover materials in the active landfill. The monetary savings can, thus, be a strong stimulus to policy to shift to sustainable ideas of material reuse and recycling.

Research gap

In Palestine, there is a research gap in the urban mining of compost and its potential contribution to the circular economy. Although many studies were conducted on compost production and uses, none of these studies evaluated the value chain of compost production and utilization to highlight its role in the circular economy and sustainability. This is the first study in Palestine that considers the current situations including the waste management and agricultural sectors' problems and highlight the importance of organic waste mining, composting, and its role in the circular economy and contribution to solve such problems. On the policy level, there are no guidelines regarding urban and landfill mining of solid waste in Palestine in general and in the study area in particular. Although there is a severe limitation on land use for landfills construction and expansion due to political reasons, there is no obligatory policy or legislative framework for waste reuse and recycling as well, and landfill mining has never been mentioned. Therefore, this study paves the way toward setting policy for urban mining of municipal SW to enhance circular economy in waste management.

Materials and methods

The study area

The study area is located in the southern part of the West Bank and consists of two governorates according to the governmental managerial system. These governorates are Hebron and Bethlehem, with a total population of 922,064 [52] and an annual growth rate of 2.8% [53].

Generation of solid waste

In accordance with the unpublished data obtained from the JSC-H&B, per capita waste generation was 0.69 kg/d and 0.79 kg/d in the Hebron and Bethlehem governorates respectively [32]. Currently, the study area is generating 925 tons daily as per the records of the JSC-H&B weighing bridge [54]. The waste characteristics are organics 46.0%, paper and cardboard 10.9%, plastic containers 18.3%, textile 6.1, metals 1.8%, glass 2.3%, and others 14.6% [32]. The highest fraction is organic which equals 425.5 tons per day.

Use of compost and chemical fertilizers in agriculture

The benefits of compost for agricultural purposes are widely reported in the literature. It can be used for soil bioremediation [55], control plant disease [56], weed control [57], improve soil biodiversity [58], improve soil fertility and water holding capacity [59], and substitute chemical fertilizers [60–62]. Palestinian farmers' acceptance of the use of compost in agriculture has been assessed at different locations. An assessment of the farmers' attitudes toward the use of compost in agriculture in Wadi Al-Far'ah Watershed in the north of the West Bank of Palestine [63] found that 84% of the farmers have the willingness to produce and use compost in agriculture. A study conducted in the Hebron district in the study area showed the farmers' attitudes toward the use of compost in agriculture are positive as 91.2% of the served farmers agree to use the compost [64].

The demand for compost in Palestine is increasing due to the shortage of chemical fertilizers in the market, because the Israeli occupation has restricted imports of dual-use fertilizers and chemicals [65]. These include fertilizers like

ammonium nitrate (NH_4NO_3), potassium nitrate (KNO_3), urea (CH_4N_2O), urea nitrate ($CH_4N_2ONO_3$), fertilizers containing nitrogen, phosphorus, and potassium (at percentages of 27-10-17 and 20-20-20) and any other fertilizer containing any ammonium nitrate, potassium nitrate, or urea [65, 66]. This has forced the Palestinian farmers to use alternative fertilizers with lower nutrient concentrations which is inappropriate and sometimes ineffective. The use of such low nutrient concentration fertilizers heavily burdened the farmers and incurred high costs due to the relatively large quantity of fertilizers they use to compensate for the low concentration, which in turn increases the soil salinity and deterioration of the agricultural productivity. It has been reported that the agricultural productivity declined in the range of 20–33% due to the use of N, P, K chemical fertilizers type 13-13-13 rating instead of 20-20-20 rating, and the farmers have borne extra costs of $29 million/year as well [66]. Under the current conditions of such severe restrictions on the import of chemical fertilizers, compost is the best alternative that can remediate the soil after the intensive use of inappropriate chemical fertilizers in relatively large quantities.

Use of compost as landfill cover

One of the uses of compost is replacing filling materials as landfill cover. This use is also widely reported in the literature [67–69]. Applying compost as cover materials over the waste at the landfills is considered an environmental benefit that can mitigate, the spread of odors, flies, mosquitoes, and reduce the release of methane. Methanotrophs in compost can oxidize methane gas and convert it into carbone dioxide, water, and biomass under aerobic conditions [34–36]. The introduction of compost as landfill cover could be one of the viable options to enhance methane reduction in landfills [70]. In addition, it can achieve financial benefits by reducing the preparation and transportation of filling materials to cover the landfill. A study on compost extracted from landfill mining [71] reported that the compost quality parameters exceed the quality limits of fertilizers, and accordingly, the extracted compost is suitable only for alternative landfill cover.

In the study area, the current cost of each ton of soil used as cover material is NIS 13 (USD 4.0/ton) [72]. Therefore, the diversion of organic waste from the SW stream and using it as cover materials after composting is a viable and sustainable option and supports the circular economy.

Methodology

A proposed model for waste management in the study area including the diversion of the organic fraction for recycling as an indicator for the circular economy is presented in Fig. 1. The saving will be revenues generated from the compost selling, saving the dumping fees that is normally paid for waste dumping at the landfill, saving from the methane emissions reduction, and saving from the reduction of the use of chemical fertilizers.

In Palestine, the landfill fee is USD 9.5/ton (NIS 30) which is usually paid by the local authority to the JSC-H&B for waste dumping. Methane emission reduction is paid through the Carbon Credit–Clean Development Mechanism (CDM) and equals USD 23.2/ton CO_{2eq} as per [73], although the CDM is not functioning in Palestine at the present. The solid waste management sector is contributing by 23% of the total emissions in Palestine [48]. The amount of methane gas emission saving that is supposed to be emitted from the waste diverted from the waste stream is calculated using the landfill gas emission model (LandGEM) version 3.03 designed by the United States Environmental Protection Agency. However, compost produced from organic waste is replacing synthetic fertilizers; therefore, the saving is the revenues generated from selling the compost and the saving from the replacement of the synthetic fertilizers.

The selling price of the compost as provided by a study conducted by the IFC on municipal organic waste composting in Palestine is USD 33/ton (NIS 100/ton) [32] considering the quality and promotion of the users. The saving in chemical fertilizers is estimated based on the nutrients which include carbon (C), Nitrogen (N), Potassium (K), and Phosphorus (P), which are available in chemical fertilizers and supplied to the plants. The commercial source of these nutrients is available in the form of Iperen Humic 12 + 3 liquid for C, Ammonium Sulphate for N, Triple Super phosphate (TSP) for P, and potassium sulphate (SOP) for K. The nutrient content in compost as reported by [74, 75] is 21.02% for C, 0.935% for N, 0.32% for K, and 0.235% for P, which will replace that available in the synthetic fertilizers when using the compost in agriculture. The nutrient content in the synthetic fertilizers and their prices are: Iperen Humic 12 + 3 liquid contains 7.3% C [76] with a price of USD 3,750.0/ton (NIS 12,000.0/ton), Ammonium Sulphate (AS) 21% N [77] with a price of USD 1,250.0/ton (NIS 4,000.0/ton), TSP contains 46% P [78] in the form of P_2O_5 with a price of USD 1,406.0/ton (NIS 4,500.0/ton), and SOP contains 50% K [79] with a price of USD 1,500.0/ton (NIS 4,800.0/ton). The prices of the fertilizers are provided by Sharabati Brothers Trading and Agricultural Company (SBTAC) [80].

The saving in the landfill volume is also considered depending on the construction cost of USD 2.64/m^3 space [81], and the waste density at the landfill after compaction which reached 0.93 [82]. Using the density in place, the space occupied by 1 ton of organic waste costs USD 2.84 (NIS 9.0/ton). The waste collection cost is excluded from the calculations, because the composting facilities is on the same site as the landfill, and in normal operation,

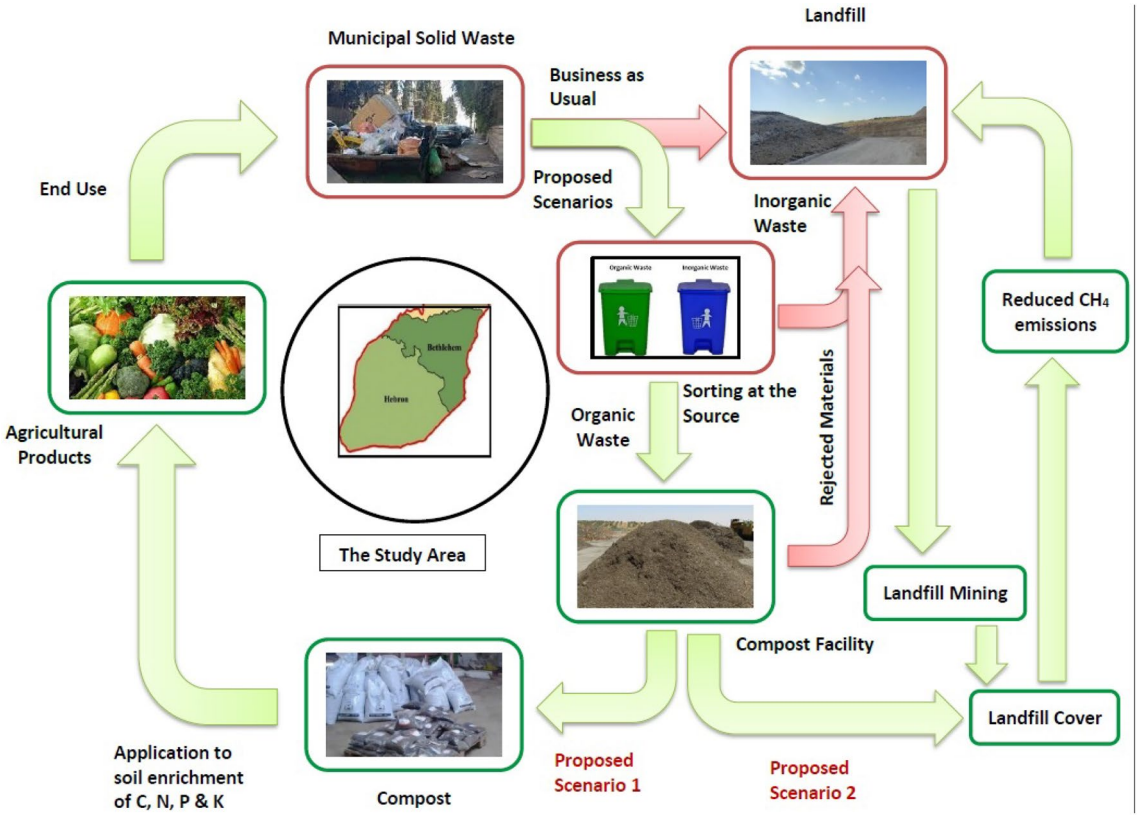


Fig. 1 Proposed waste management model for Hebron and Bethlehem governorates, Palestine

the waste is collected and brought to the site, so there is no difference in the waste collection cost between both scenarios, landfilling and composting.

The saving is calculated based on the following:

$$\mathrm{Rn} = \mathrm{Rg} - \mathrm{OM} \tag{1}$$

$$\mathrm{Rg} = \mathrm{D} + \mathrm{Cc} + \mathrm{C} + \mathrm{F} + \mathrm{Ls} \tag{2}$$

where Rn is the net revenue; Rg is the gross revenue; OM is the operation and maintenance cost of the composting plant; D is the dumping fee saving; Cc is the carbon credit value; C is the compost selling; F is the fertilizer savings; and Ls is the landfill space saving.

The trend is taken through the coming 15 years starting from 2021 to 2035. To calculate the waste generation projection, the population is projected using a 2.8% growth rate [53], and population baseline in accordance with the census in 2017 [52], and waste projection is calculated in accordance with Eq. 3 which is derived based on the actual SW quantities generated in 2015–2020.

$$\mathrm{WasteQuantity} = 850,744.61\ln(x) - 11,409,010.59 \tag{3}$$

where x is the number of population.

The organic compostable fraction is calculated based on the percentage of the organic waste fraction from the total waste stream based on the waste characterization which is 46%. It has been assumed that only 70% of the organic waste fraction is compostable [32] as a conservative figure taking into account that the waste separation at the source could be insufficient, which results in some rejected materials. The waste projection is shown in Fig. 2.

For compost yield, it has been assumed that 25% of organic waste is retrievable with 40% moisture content [74, 75].

This methodology can be used for other similar types of materials and wastes depending on the option of reuse, recyclability, or upscaling of refused material. The materials like paper, plastic, and metal have got fair scrap value recyclability in most of the countries [83] resulting in the generation of resource and revenue from the waste. Efforts are underway to effectively utilize the construction and Demolition waste [84] and agro-residues [85] to overcome the environmental effects and support circular economy. Since the present study integrates the waste utilization and economic growth while adopting nature-based solutions to overcome major issues related to environment and public health, it shall serve

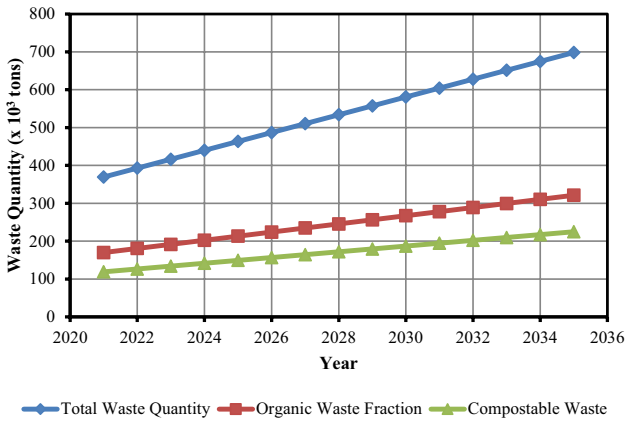


Fig. 2 Projection of waste generation in the study area from 2021 to 2035

as an important document for making rational decisions and framing policies.

Results and discussion

Based on the objectives of the present study, the results have been classified into two scenarios—the use of compost in agriculture, and the use of compost as landfill cover, as given below.

Scenario 1: use of compost in agriculture

Compost production and selling revenues

The organic waste generation is about 118,948.29 tons in 2021 and will reach 224,856.78 tons in 2035. It has been assumed that the organic waste generated will yield 25% compost, so the compost production is in the range of 29,737–56,214 tons for 2021–2035. The operation and maintenance (O&M) cost of the compost facility is assumed to be 60% of the compost sales considering the low prices of compost. Other researchers assumed the O&M cost to be 30% of the compost selling prices for operating and maintaining a compost facility [75]. However, the revenues generated from compost selling are in the range of USD 392,529–742,027 from 2021 to 2035. Table 1 presents the revenues for each year.

Potential replacement of chemical fertilizers and monetary saving

Compost can be used for agricultural purposes, and once it is produced and used, it enters the cycle of the circular economy. The available chemical fertilizers in Palestine and mostly used in agriculture are Ammonium Sulphate (AS), Triple Super Phosphate (TSP), and Potassium Phosphate (SOP). The calculations showed that 17,842 tons of compost can replace 794 tons of AS, 91 tons of TSP, and 29 tons of SOP in 2021, and 33,729 tons of compost can replace 1,502 tons of AS, 172 tons of TSP, and 54 tons of SOP in 2035, as shown in Fig. 3 and Table 2.

A study conducted in the Kingdom of Saudi Arabia (KSA) [75] using commercial fertilizers in the KSA found

Table 1 Compost production and revenues during the period 2021–2035

Year	Waste quantity (ton)	Compost (ton)	Selling value (USD)	O&M cost (USD)	Revenue (USD)
2021	118,948	29,737	981,323	588,794	392,529
2022	126,513	31,628	1,043,734	626,240	417,494
2023	134,078	33,520	1,106,144	663,687	442,457
2024	141,643	35,411	1,168,555	701,133	467,422
2025	149,208	37,302	1,230,965	738,579	492,386
2026	156,773	39,193	1,293,375	776,025	517,350
2027	164,338	41,084	1,355,786	813,471	542,314
2028	171,903	42,976	1,418,196	850,918	567,278
2029	179,467	44,867	1,480,606	888,364	592,243
2030	187,032	46,758	1,543,017	925,810	617,207
2031	194,597	48,649	1,605,427	963,256	642,171
2032	202,162	50,541	1,667,837	1,000,702	667,135
2033	209,727	52,432	1,730,248	1,038,149	692,099
2034	217,292	54,323	1,792,658	1,075,595	717,063
2035	224,857	56,214	1,855,069	1,113,041	742,027

Operation and maintenance cost is assumed 60% of the compost selling revenues

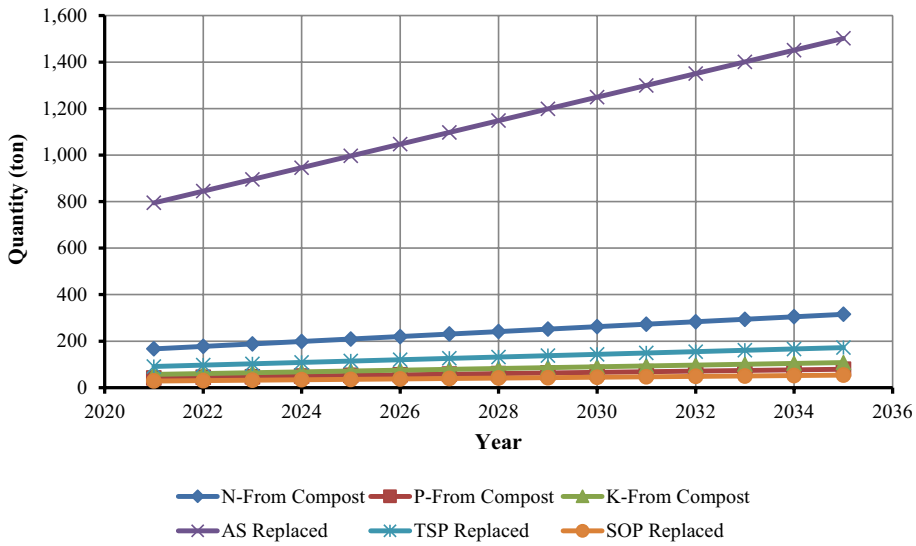


Fig. 3 Replacement of Ammonium Sulphate (AS), Triple Super Phosphate (TSP), and Potassium Phosphate (SOP) by compost

Table 2 Replacement of AS, TSP, and SOP by compost

Year	Compost (ton)	Compost DM (ton)	N, P, and K in compost (ton)			AS, TSP, and SOP replaced (ton)		
			N	P	K	AS	TSP	SOP
2021	29,737	17,842	167	42	57	794	91	29
2022	31,628	18,977	177	45	61	845	97	30
2023	33,520	20,112	188	47	64	895	103	32
2024	35,411	21,246	199	50	68	946	109	34
2025	37,302	22,381	209	53	72	996	114	36
2026	39,193	23,516	220	55	75	1,047	120	38
2027	41,084	24,651	230	58	79	1,098	126	39
2028	42,976	25,785	241	61	83	1,148	132	41
2029	44,867	26,920	252	63	86	1,199	138	43
2030	46,758	28,055	262	66	90	1,249	143	45
2031	48,649	29,190	273	69	93	1,300	149	47
2032	50,541	30,324	284	71	97	1,350	155	49
2033	52,432	31,459	294	74	101	1,401	161	50
2034	54,323	32,594	305	77	104	1,451	167	52
2035	56,214	33,729	315	79	108	1,502	172	54

DM: dry matter; compost contains 40% moisture content, 21.02% carbon, 0.935% N, 0.235% P, and 0.32% K. Alternative source of carbon such as humic acid used is "Iperen Humic 12 + 3 liquid" contains 7.3% carbon; ammonium Phosphate contains 21% N; TSP contains 46% P, and SOP contains 50% K

that 0.23 Mt of compost produced from organic food waste of Makkah city can replace 2.8, 0.7, and 0.9kt of Urea, diammonium phosphate (DAP), and sulphate of potash. This study showed that the same quantity of compost replaced AS, TSP, and SOP can enrich the soil with 3,750 tons of C which in turn can replace 51,376 tons of alternative fertilizers of humic acid (commercially used is Iperen Humic 12 + 3 liquid which contains 7.3% C) in 2021. The replacement increases with time as the compost quantity increases based on the waste generation. The projection of replacement showed that 33,729 tons of compost can enrich the soil with 7,090 tons of C which can replace 97,120 tons of "Iperen Humic 12 + 3 liquid" in 2035 as presented in Table 3. Previous research [66] found that 28.9 and 50.4 kt of C from compost quantities 0.23 and 0.40 Mt can replace 119.4 and 208.8 kt of humic acid in 2015 and 2030, respectively. Moreover, compost builds up organic matter which confers long-term nutrient release (approximately 6 years) [86], which is challenging to estimate. Taking this long-term effect of compost on the soil, the substitution value would be much higher during the subsequent years [87].

Table 3 Compost soil enrichment with C and replacement of alternative fertilizers

Year	Compost (ton)	Compost DM (ton)	C-enrichment (tons)	Alternative source (tons)
2021	29,737	17,842	3,750	51,376
2022	31,628	18,977	3,989	54,643
2023	33,520	20,112	4,228	57,911
2024	35,411	21,246	4,466	61,178
2025	37,302	22,381	4,705	64,446
2026	39,193	23,516	4,943	67,713
2027	41,084	24,651	5,182	70,980
2028	42,976	25,785	5,420	74,248
2029	44,867	26,920	5,659	77,515
2030	46,758	28,055	5,897	80,783
2031	48,649	29,190	6,136	84,050
2032	50,541	30,324	6,374	87,317
2033	52,432	31,459	6,613	90,585
2034	54,323	32,594	6,851	93,852
2035	56,214	33,729	7,090	97,120

Compost contains 40% moisture and 21.02% carbon, alternative source of carbon such as humic acid used is "Iperen Humic 12 + 3 liquid" which contains 7.3% carbon; ammonium Phosphate contains 21% N; TSP contains 46% P, and SOP contains 50% K

The monetary value of the saving from the replacement of the chemical fertilizers is depicted in Fig. 4. The estimated results show that the saving generated is USD 993,005.80, USD 128,157.73, and USD 42,821.38 from AS, TSP, and SOP, respectively, in 2021. The saving rises as the quantity of compost increased with time based on the population growth and reached USD 1,877,152.60 from AS USD 242,266.07 from TSP, and USD 80,948.44 from SOP in 2035. In addition, the estimated saving from replacing "Iperen Humic 12 + 3 liquid", the source of C, is USD 192.7 million in 2021 and USD 364.2 million in 2035. The total estimation of C, N, P, and K replacement savings of "Iperen Humic 12 + 3 liquid", AS, TSP, and SOP by compost is USD 193.8 million and USD 366.4 million in 2021 and 2035, respectively.

Reduction in methane emission and monetary saving

Reduction in methane emission is considered an environmental indicator of the circular economy. In the proposed scenario, organic waste will be composted and methane emissions will be prevented. The calculated results show that the methane emissions reduction is 165 tons, which is equal to 4,122 tons of CO_2 reduction in 2022, while the LGEM does not provide estimates in the first year (2021) and provides estimates next year from the start. By the year 2035, the methane emission reduction will reach 2,493 tons which is equivalent to 62,324 of CO_2 which is prevented from being released to the atmosphere, thus reducing the potential global warming and mitigating climate change. In a similar study, [66] found that composting of food waste can reduce CO_2eq emissions by 1.09 and 1.9 Mt in 2015 and 2030, respectively. The monetary value of this reduction according to the CDM is presented in Table 4.

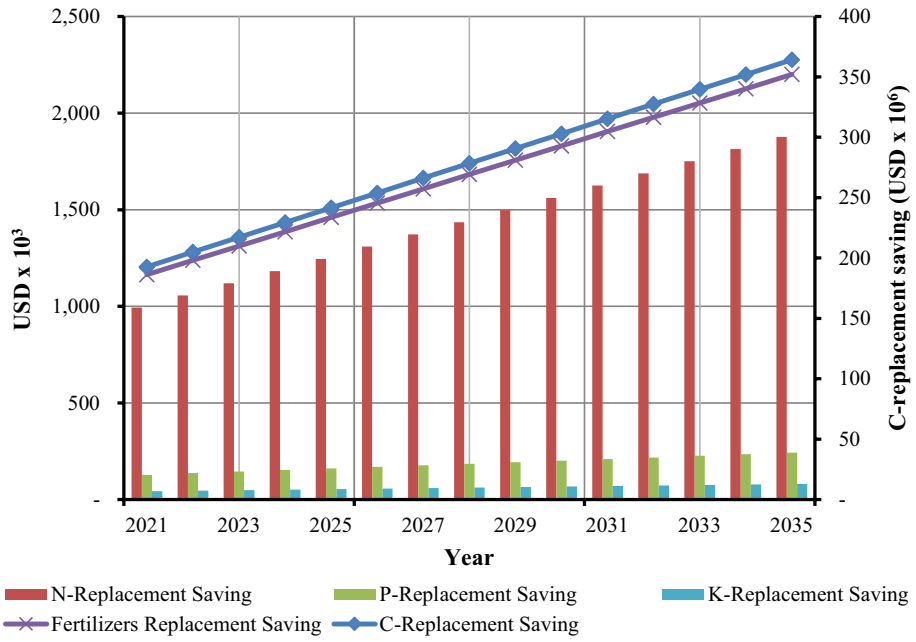


Fig. 4 Saving from replacement of chemical fertilizers by compost in Palestine

Table 4 Saving from CDM and dumping fees

Year	CH_4 (ton)	GWP CO_2 eq (tons)	Saving from CDM (USD)	Dumping fees (USD)	Total saving (USD)
2021	0	0	0	282,502	282,502
2022	165	4,122	95,640	300,469	396,109
2023	332	8,306	192,699	318,435	511,134
2024	502	12,548	291,106	336,402	627,508
2025	674	16,845	390,796	354,369	745,165
2026	848	21,194	491,708	372,335	864,043
2027	1,024	25,594	593,780	390,302	984,082
2028	1,202	30,041	696,956	408,269	1,105,225
2029	1,381	34,534	801,183	426,235	1,227,418
2030	1,563	39,069	906,410	444,202	1,350,612
2031	1,746	43,646	1,012,587	462,168	1,474,755
2032	1,931	48,262	1,119,668	480,135	1,599,803
2033	2,117	52,914	1,227,609	498,102	1,725,711
2034	2,304	57,602	1,336,369	516,068	1,852,437
2035	2,493	62,324	1,445,907	534,035	1,979,942

GWP of CH_4 = 25 CO_2; CO_2 saving price = USD 23.2/ton; dumping fees = USD 9.5/ton

Saving from dumping fees

The tipping fee for municipal waste at the regional landfill is USD 9.5/ton (NIS 30/ton). As the organic waste will be diverted from the waste stream, this fee will not be paid and is considered as saving. The calculations presented in Table 3 showed that the saving in 2021 from the dumping fees is USD 282,502. However, this saving will reach USD 534,035 in 2035. Similar to the present study, UCTAD [66] reported that composting food waste can achieve saving in the tipping fees of 524.25 and 914.19 million Saudi Riyal in 2015 and 2030, respectively.

Environmental saving

Saving achieved from the tipping fees and from the reduction of methane emissions in accordance with the CDM is considered *environmental saving,* and also is one of the indicators of the circular economy in organic waste management and composting. Through CDM, developing countries can implement projects to reduce carbon emissions, and this reduction can be sold to industrial countries to enable them to meet their targets according to Kyoto Protocol. Diversion of organic waste from the waste stream and composting of this waste will, definitely, reduce carbon emissions, where this reduction can be sold. The return generated from selling carbon is considered as an environmental saving. The calculations presented in Table 3 show that the environmental saving (total saving) is USD 282,502.19 and USD 1,979,941.67 in 2021 and 2035, respectively.

Landfill space saving

Diverting organic waste from the municipal solid waste stream causes a significant reduction in waste quantities and volumes that is currently sent to the landfill for disposal. This creates considerable savings in the landfill volume occupied by this waste stream. The landfill space saved from organic waste diversion can be used for other waste streams, thus increasing the life of the landfill and reducing the costs required for expanding the landfill or constructing a new landfill. As these costs are drawn from the budget, then saving the landfill space can reduce expenditure and support the treasury as well, which is considered an indicator of the circular economy. The landfill space construction cost USD 2.65/m^3 [81], which equals USD 2.84/ton of waste considering the density of the waste after compaction at the landfill 0.93 [82]. The calculation results of the landfill space saving are USD 337,813 in 2021 and USD 638,593 in 2035 as presented in Fig. 5.

Net revenues

The net revenue is calculated based on Eqs. 1 and 2 and presented in Fig. 6. It has been estimated that a sustainable composting system can add net revenue of USD 194.8 million to the national economy in 2021 and USD 369.8 million in 2035. Saving from the replacement of chemical fertilizers is relatively high, because the prices of chemical fertilizers in Palestine are very high compared to other countries where there is free access to fertilizers and the governments support the agricultural sector with fertilizers of viable prices. For example, the price of an alternative (C) source in Saudi

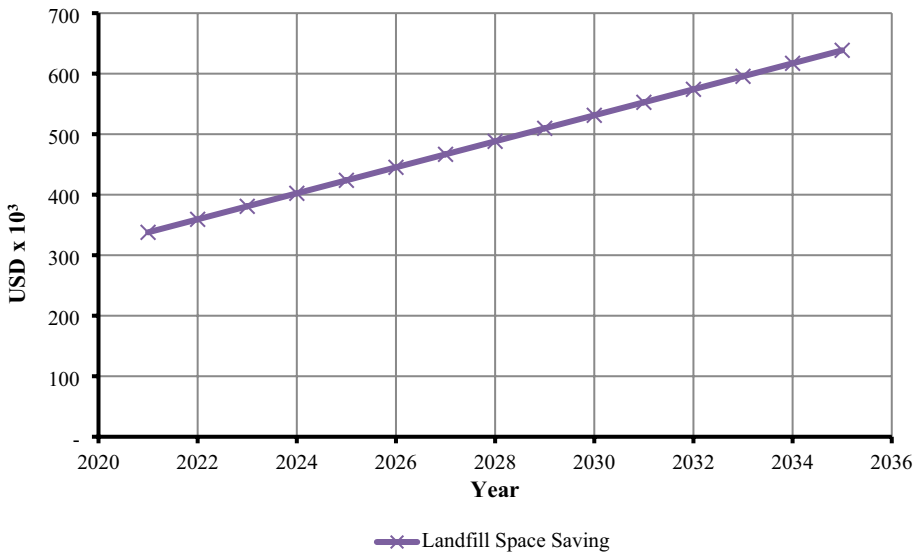


Fig. 5 Saving with respect to reduced space requirement for landfilling

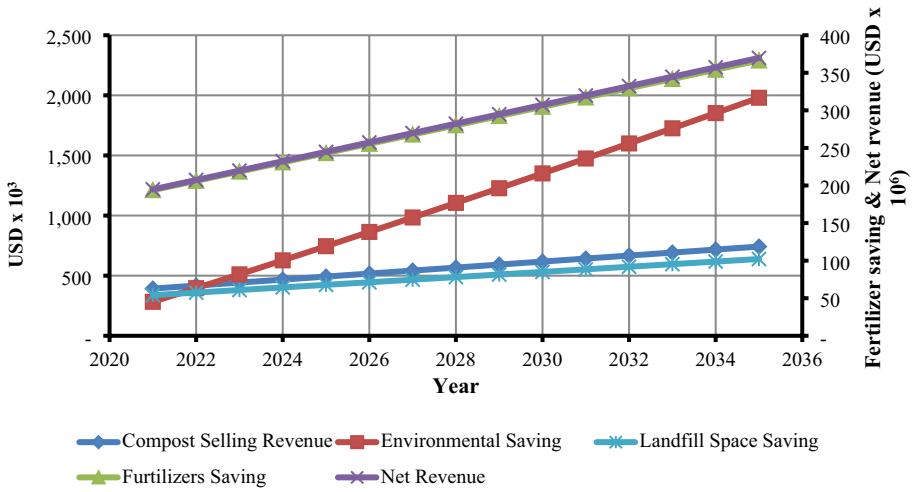


Fig. 6 Net revenues generated from compost selling, environmental saving, fertilizer replacement saving and landfill space saving

Arabia is 562.5 SAR/ton [75] which is equal to USD 150/ton. A study conducted in Kampala/Uganda on the recovery of fecal sludge, sewage sludge, and organic fraction of municipal solid waste, and found that the recovery through composting can generate USD 5.1 million [88].

Scenario 2: use of compost as landfill cover

Besides the agricultural uses of compost, it can also use as daily and intermediate landfill cover, thus saving the cost of the filling materials that are used on daily basis in the business-as-usual scenario. Landfill daily cover is essential, because it prevents the spread of flies, mosquitoes, and insects, and reduces odor emission. At the regional landfill in the study area, about 600 tons of filling materials consisted of crushed stone, marl, sand, and clay is used to cover the daily cell and working face. The price of the cover materials is USD 4.0 (NIS 13.0) per ton of cover material [72]. Accordingly, the cost of this quantity is USD 2,400.0/day. The landfill is open for 12 h per day and all days of the year. The annual estimated cost of the cover materials is USD 876,000.0 and USD 13.14 million during the 15 years starting from 2021 and up to 2035. These amounts can be saved if the compost is used as an alternative cover. Saving this amount is supporting the concept of the circular economy as the waste is recycled and returns to the material cycle and achieves financial profits. This scenario is less attractive compared with scenario 1 (use of compost in agriculture) based on the revenues generated. This agrees with Sardarmehni et al. [89] who modeled the use of compost as a soil amendment and alternative daily cover based on environmental impact and found that compost use as a soil amendment is preferable.

Impact on the long-term retention phase

The benefits of the organic waste composting increase on the long-term retention phase, and can contribute to sustainable waste management and agriculture. Given the restrictions on land use for landfill construction and expansion, and restricted access to chemical fertilizers in Palestine [65], composting can contribute to solving the current technical and financial problems of waste management and the agricultural sector, so it can reduce the demand for landfill space required for waste landfilling, thus, reducing the demand on the construction of new landfills or extension of the existing ones, and farmers will have access to compost to replace the low-quality chemical fertilizers currently in use [66]. In addition, it can reduce the demand for water for agriculture as compost can improve the water holding capacity and increase moisture retention time in the soil as well [59], knowing that water shortage is one of the biggest problems being faced in the agricultural sector in Palestine [90]. Therefore, compost can contribute to solving the problems related to fertilizers and water, thereby supporting food security in the country which suffers from political instability due to the Israeli–Palestinian conflict. However, agencies responsible for SWM can use the revenues as subsidies to cover part of the cost associated with SWM which in turn can improve the service. The environmental benefits are invaluable and are beneficial to humans, the climate, the soil, air quality and health, the water, and society, as well. It can support the sustainability pillars as well as contribute to achieving the sustainable development goals in Palestine at large.

However, a matter of concern is the reduction in the agricultural yield in the case of substitution of synthetic fertilizers by compost, because the quality of compost depends primarily on the raw materials. Therefore, the quantity of compost to be added depends on the nutrient available in the compost and the crop nutrient demand. Opinions about organic farming and inorganic farming are widely varied. Some studies consider organic farming is insufficient to meet the food demand [91], some see great potential of organic farming to secure food in an environmentally friendly manner [92], and others recommended the combined use of compost and chemical fertilizers [93, 94]. The later requires further studies considering the crop pattern to optimize the mixing ratio of compos and chemical fertilizers. However, sustainable agriculture takes into account the sustainability pillars: economic benefit, social acceptance, and environment protection. From sustainability point of view, the replacement of chemical fertilizers by compost in agriculture can reduce the crop yield [95], but ensures environmental protection and food safety.

Further, substitution of chemical fertilizers by compost depends on the country's internal situations and the global context. In Palestine, under the shortage of water for irrigation, restrictions on the import of good quality fertilizers as mentioned above, and a wide group of farmers depend on rain-fed agriculture, substitution of synthetic fertilizer by compost is a favorable option. Heydarzadeh et al. [96] recommended the substitution of chemical fertilizers by vermicompost and poultry manure in rain-fed agriculture. Moreover, the global market prices of synthetic fertilizers escalated by fourfold [97] due to Russia–Ukraine conflict, which makes the affordability of synthetic fertilizers under question especially in low- and middle-income countries.

Conclusion

The municipal SW in Palestine has the potential for recycling the organic fraction through composting and returns it to the material cycle as a resource for production as part of the circular economy. This study provides the economic and environmental benefits of compost use for agricultural purposes as well as cover materials in landfill. The estimated net revenue for the agricultural use of compost is USD 194.8 million in 2021 and USD 369.8 million in 2035 which supports the national economy and is a type of circular economy. However, using compost as landfill cover can achieve financial savings through a reduction in the use of the soil as cover materials, which is estimated annually at USD 0.876 million and USD 13.14 million throughout the study period. The results of this study can pave the way toward framing national policies and implementation of the circular economy in SWM in Palestine.

Novelty of the study

It is the first study in Palestine that focuses on nutrient recovery from organic municipal solid waste through a circular economy approach, which shed the light on recycling and could direct future studies to focus on other SW fractions following the same approach. Moreover, the study provides a solution for three major problems, including problems related to landfill construction and expansion, problems related to the restriction forced by the occupation on import of good quality fertilizers for agricultural purposes, and environmental problems related to improper waste management.

Limitations of the study

This study focused only on urban mining of the organic fraction of the municipal SW and its role in the circular economy through composting, which could open the door for future studies to follow the same circular economy approach and focus on other municipal SW fractions, such as metals,

plastics, WEEE, and glass. Although the study highlighted the potential of compost production from landfill mining, its use in agriculture and mineral recovery calculations was excluded from this study due to the absence of data on the quantities of the waste deposited in old landfills, in addition to high uncertainties of the quality due to potential heavy metals' contamination and safety concerns as well. The waste management model suggested waste separation at the source to avoid quality problems for using compost in agriculture, and use compost from landfill mining as a cover material for active landfill only.

Author contributions MIAS: writing—review and editing, conceptualization, methodology, data curation and calculation, writing—original draft, and writing—review and editing. AKH: supervision, and writing—review and editing.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest None.

References

1. Mayer A, Haas W, Wiedenhofer D, Krausmann F, Nuss P, Blengini GA (2019) Measuring progress towards a circular economy: a monitoring framework for economy-wide material loop closing in the EU28. J Ind Ecol 23(1):62–76. https://doi.org/10.1111/jiec.12809
2. EU (European Commission) (2020) Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A New Circular Economy Action Plan for a Cleaner and More Competitive Europe, Communication No. 98; European Commission: Brussels, Belgium. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2020%3A98%3AFIN. Accessed on 7 Jan 2022.
3. K Flanagan, A Clowes, B Lipinski, L Goodwin, R Swannell (2018) SDG Target 12.3 on food loss and waste: 2018 progress report. An annual update on behalf of Champions 12.3. https://champions123.org/sites/default/files/2020-09/champions-12-3-2018-progress-report.pdf. Accessed on 07 Jan 2022.
4. Ellen Macarthur Foundation (2015) Toward a circular economy: business rationale for an accepted transition. https://emf.thirdlight.com/link/ip2fh05h21it-6nvypm/@/preview/1?o. Accessed on 20 Nov 2021.
5. Cossu R, Williams ID (2015) Urban mining: concepts, terminology, challenges. Waste Manage 45:1–3. https://doi.org/10.1016/j.wasman.2015.09.040
6. Sehnem S, Ndubisi NO, Preschlak D, Bernardy RJ, Santos Junior S (2020) Circular economy in the wine chain production: maturity, challenges, and lessons from an emerging economy perspective. Prod Plan Control 31(11–12):1014–1034. https://doi.org/10.1080/09537287.2019.1695914
7. Slorach PC, Jeswani HK, Cu´ellar-Franca R, Azapagic A (2020) Assessing the economic and environmental sustainability of household food waste management in the UK: current situation and future scenarios. Sci Total Environ 710:135580. https://doi.org/10.1016/j.scitotenv.2019.135580
8. Razza F, D'Avino L, L'Abate G, Lazzeri L (2018) The Role of Compost in Bio-waste Management and Circular Economy. Research Gate. In: Benetto E, Gericke K, Guiton M (eds) Designing Sustainable Technologies Products and Policies: From Science to Innovation. Springer International Publishing, Cham, pp 133–143
9. ECN (European Compost Network) (2016). Bio-waste: the valuable organic resource in a circular economy. Fact sheet. https://www.compostnetwork.info/ecns-factsheet-organic-resource/. Accessed on 20 Nov 2021.
10. S Rau (2019) Options for Urban Mining and Integration with a Potential Green Circular Economy in the People's Republic of China. ADB BRIEFS 124. https://doi.org/10.22617/BRF190606-2.
11. Park JK, Mahoney J, Clark T, Krueger N (2017) A review of urban mining in the past, present and future. Adv Recycling Waste Manag 2:127. https://doi.org/10.4172/2475-7675.1000127
12. Baccini P, Brunner P (2012) Metabolism of the Antroposphere: Analysis, Evaluation, Design, 2nd edn. The MIT Press, Cambridge, MA. https://doi.org/10.7551/mitpress/8720.001.0001
13. Arora R, Paterok K, Banerjee A, Saluja MS (2017) Potential and relevance of urban mining in the context of sustainable cities. IIMB Manag Rev 29:210–224. https://doi.org/10.1016/j.iimb.2017.06.001
14. Tejaswini MSSR, Pathak P, Gupta DK (2022) Sustainable approach for valorization of solid wastes as a secondary resource through urban mining. J Environ Manage 319:115727. https://doi.org/10.1016/j.jenvman.2022.115727
15. Chand S, Ranjan Rout P, Pathak P (2021) Basic Concepts, Potentials, and Challenges of Urban Mining Urban Mining for Waste Management and Resource Recovery: Sustainable Approaches. CRC Press, Taylor & Francis Group, Boca Raton, FL, USA, pp 1–11
16. Schlezak SL, Styer JE (2023) Inclusive urban mining: an opportunity for engineering education. Mining 3:284–303. https://doi.org/10.3390/mining3020018
17. Ghisellini P, Ncube A, Casazza M, Passaro R (2022) Toward circular and socially just urban mining in global societies and cities: present state and future perspectives. Front Sustain Cities 4:1–14
18. Ntapanta SM (2021) 'Lifescaping' toxicants: locating and living with e-waste in Tanzania. Anthropol Today 37:7–10
19. Xavier LH, Giese EC, Ribeiro-Duthie AC, Lins FAF (2021) Sustainability and the circular economy: a theoretical approach focused on e-waste urban mining. Res Policy 74:101467. https://doi.org/10.1016/j.resourpol.2019.101467
20. Zheng X, Xiao T, Xu G, Albalghiti E, Shan G, Li J (2021) Comparing the costs and benefits of virgin and urban mining. J Manage Sci Eng 7:98–106. https://doi.org/10.1016/j.jmse.2021.05.002
21. Ghisellini P, Ulgiati S (2020) Circular economy transition in Italy. Achievements, perspectives and constraints. J Clean Prod 243:118360
22. Girotto F, Alibardi L, Cossu R (2015) Food waste generation and industrial uses: a review. Waste Manage 45:32–45. https://doi.org/10.1016/j.wasman.2015.06.008
23. Ddiba D, Andersson K, Rosemarin A et al (2022) The circular economy potential of urban organic waste streams in low- and middle-income countries. Environ Dev Sustain 24:1116–1144. https://doi.org/10.1007/s10668-021-01487-w
24. Awasthi SK, Sarsaiya S, Awasthi MK, Liu T, Zhao J, Kumar S, Zhang Z (2020) Changes in global trends in food waste composting: research challenges and opportunities. Bioresour Technol 299:122555. https://doi.org/10.1016/j.biortech.2019.122555
25. Chang CT, Negi S, Rani A, Hu AH, Pan SU, Kumar S (2022) Food waste and soybean curd residue composting by black soldier

fly. Environ Res 214:113792. https://doi.org/10.1016/j.envres.2022.113792
26. Duan Y, Awasthi SK, Liu T, Pandey A, Zhang Z, Kumar S, Mk A (2020) Succession of keratin-degrading bacteria and associated health risks during pig manure composting. J Clean Prod 258:120624. https://doi.org/10.1016/j.jclepro.2020.120624
27. Geng Y, Cao G, Wang L, Wang S (2019) Effects of equal chemical fertilizer substitutions with organic manure on yield, dry matter, and nitrogen uptake of spring maize and soil nitrogen distribution. PLoS ONE 14(7):e0219512
28. Li J, Otero-Gonzalez L, Parao A, Tack P, Folens K, Ferrer I, Lens PNL, Laing GD (2021) Valorization of selenium-enriched sludge and duckweed generated from wastewater as micronutrient biofertilizer. Chemosphere 281:130767
29. Wang Z, Ding Y, Ren X, Xie J, Kumar S, Zhang Z, Wang Q (2022) Effect of micronutrient selenium on greenhouse gas emissions and related functional genes during goat manure composting. Bioresour Technol 349:126805. https://doi.org/10.1016/j.biortech.2022.126805
30. Cui X, Lu H, Lu Y, Gao P, Peng F (2022) Replacing 30% chemical fertilizer with organic fertilizer increases the fertilizer efficiency, yield and quality of cabbage in intensive open-field production. Cienc Rural 52(7):e20210186. https://doi.org/10.1590/0103-8478cr20210186
31. Esmaeilian Y, Amiri MB, Tavassoli A, Caballero-Calvo A, Rodrigo-Comino J (2022) Replacing chemical fertilizers with organic and biological ones in transition to organic farming systems in saffron (Crocus sativus) cultivation. Chemosphere 307(1):135537. https://doi.org/10.1016/j.chemosphere.2022.135537
32. IFC (International Finance Cooperation) (2012) Solid waste management in Hebron and Bethlehem Governorates, assessment of current situation and analysis of new system. Unpublished Report, Palestine
33. Zaulfikar SB (2020) Landfill mining dominated by organic solid waste: a review on its benefits, potential and challenges to recovery landfills in growing cities in Indonesia. IOP Conf Series: Mater Sci Eng 845:012052. https://doi.org/10.1088/1757-899X/845/1/012052
34. USEPA (United States Environmental Protection Agency) (2002) Cover up with compost, Fact sheet. https://archive.epa.gov/epawaste/nonhaz/municipal/web/pdf/f02022.pdf. Accessed on 19 Nov 2021.
35. Barlaz MA, Green RB, Chanton JP, Goldsmith CD, Hater GR (2004) Evaluation of a biologically active cover for mitigation of landfill gas emissions. Environ Sci Technol 38(18):4891–4899. https://doi.org/10.1021/es049605b
36. Albanna M, Fernandes L, Mostafa W (2007) Methane oxidation in landfill cover soil; the combined effects of moisture content, nutrient addition, and cover thickness. J Environ Eng Sci 6(2):191–201. https://doi.org/10.1139/s06-047
37. Zaulfikar A, Sudarno U, Budiyono B (2020) Analysis of optimum garbage heaps age on recovery of landfills dominated by organic solid waste. J Ecol Eng 21(8):91–98
38. Prechthai T, Padmasri M, Visvanathan C (2008) Quality assessment of mined MSW from an open dumpsite for recycling potential. Resour Conserv Recy 53:70–78. https://doi.org/10.1016/j.resconrec.2008.09.002
39. Zhou C, Xu W, Gong Z, Fang W, Cao A (2015) Characteristics and fertilizer effects of soil-like materials from landfill mining. Clean: Soil, Air, Water 43:940–947. https://doi.org/10.1002/clen.201400510.487
40. Kurian J, Esakku S, Palanivelu K, Selvam A (2003) Studies on landfill mining at solid waste dumpsites in India. Sardinia, Itlay. Proc Ninth Int Waste Manage Landfill Symp 3:248–255
41. Bhatnagar A, Kaczala F, Burlakovs J, Kriipsalu M, Hogland M, Hogland W (2017) Hunting for valuables from landfills and assessing their market opportunities: a case study with Kudjape landfill in Estonia. Waste Manage Res 35(6):1–9. https://doi.org/10.1177/0734242X17697816
42. Datta M, Somani M, Ramana GV, Sreekrishnan TR (2021) Feasibility of re-using soil-like material obtained from mining of oldMSW dumps as an earth-fill and as compost. Process Saf Environ Prot 147:477–487
43. Mitra S, Chakraborty AJ, Tareq AM, Emran TB, Nainu F, Khusro A, Idris AM, Khandaker MU, Osman H, Alhumaydhi FA, Simal-Gandara J (2022) Impact of heavy metals on the environment and human health: novel therapeutic insights to counter the toxicity. J King Saud Univ Sci 34:101865
44. Coetzee JJ, Bansal N, Chirwa EMN (2020) Chromium in environment, its toxic effect from chromite-mining and ferrochrome industries, and its possible bioremediation. Expo Health 12(1):51–62. https://doi.org/10.1007/s12403-018-0284-z
45. Hou S, Yuan L, Jin P, Ding B, Qin N, Li L, Liu X, Wu Z, Zhao G, Deng Y (2013) A clinical study of the effects of lead poisoning on the intelligence and neurobehavioral abilities of children. Theor Biol Med Model 10(1):1–9. https://doi.org/10.1186/1742-4682-10-13
46. Garza-Lombó C, Pappa A, Panayiotidis MI, Gonsebatt ME, Franco R (2019) Arsenic-induced neurotoxicity: a mechanistic appraisal. J Biol Inorg Chem 24(8):1305–1316
47. Lentini P, Zanoli L, Granata A, Signorelli SS, Castellino P, Dell'Aquila R (2017) Kidney and heavy metals – the role of environmental exposure (Review). Mol Med Rep 15:3413–3419. https://doi.org/10.3892/mmr.2017.6389
48. UNDP (United Nations Development Program), (2016). Provision of services to prepare a Green House Gases (GHG) emission inventory and the mitigation chapters of Palestine's Initial National Communication Report (INCR).
49. Al-Sari' MI, Haritash AK (2024) A logistic regression model to facilitate setting of organic waste composting policy for sustainable waste management. Environ Dev Sustain. https://doi.org/10.1007/s10668-024-04934-6
50. Salah M, Al-Sari' M, Al-Khatib IA, Kontogianni S (2020) Local residents' perception of landfll impacts in Palestine: the case of Zahrat Al-Finjan landfill. J Mater Cycles Waste Manage 22:673–681
51. Al-Khatib IA, Ajlouny H, Al-Sari' MI, Kontogianni S (2014) Residents' concerns and attitudes toward solid waste management facilities in Palestine: a case study of Hebron district. Waste Manage Res 32(3):228–236
52. PCBS (Palestinian Central Bureau of Statistics) (2018) Preliminary results of the population, housing and establishments, census 2017. Ramallah, Palestine. https://www.pcbs.gov.ps/Downloads/book2364-1.pdf. Accessed on 20 Nov 2021.
53. UNPF (United Nations Population Fund) (2021) population matters. https://palestine.unfpa.org/en/population-matters-0. Accessed on 21 Oct 2021.
54. The Joint Service Council for Solid Waste Management for Hebron and Bethlehem (JSC-H&B) (2020) Weighting bridge records of solid waste quantities delivered to Al-Menya landfill. Unpublished data.
55. Ventorino V, Pascale A, Fagnano M, Adamo P, Faraco V, Rocco C, Fiorentino N, Pepe O (2019) Soil tillage and compost amendment promote bioremediation and biofertility of polluted area. J Clean Prod 239:118087. https://doi.org/10.1016/j.jclepro.2019.118087
56. Pane C, Spaccini R, Piccolo A, Celano G, Zaccardelli M (2019) Disease suppressiveness of agricultural green waste composts as related to chemical and bio-based properties shaped by different on-farm composting methods. Biol Control 137:104026. https://doi.org/10.1016/j.biocontrol.2019.104026

57. Coelho L, Osório J, Beltrão J, Reis M (2019) Organic compost effects on Stevia rebaudiana weed control and on soil properties in the Mediterranean region. Revista de Ciências Agrárias 42(1):109–121
58. Pose-Juan E, Igual JM, Sánchez-Martín MJ, Rodríguez-Cruz MS (2017) Influence of herbicide triasulfuron on soil microbial community in an unamended soil and a soil amended with organic residues. Front Microbiol 8:378. https://doi.org/10.3389/fmicb.2017.00378
59. Bekchanov M, Mirzabaev A (2018) Circular economy of composting in Sri Lanka: Opportunities and challenges for reducing waste related pollution and improving soil health. J Clean Prod 202(20):1107–1119. https://doi.org/10.1016/j.jclepro.2018.08.186
60. Rajaie M, Tavakoly AR (2016) Effect of municipal waste compost and nitrogen fertilizer on growth and mineral composition of tomato. Int J Recycl Org Waste Agric 5:339–347. https://doi.org/10.1007/s40093-016-0144-4
61. Azim K, Komenane S, Soudi B (2017) Agro-environmental assessment of composting plants in southern western of Morocco (Souss-Massa region). Int J Recycl Org Waste Agric 6:107–115. https://doi.org/10.1007/s40093-017-0157-7
62. Oliveira LSBL, Oliveira DSBL, Bezerra BS, Pereira BS, Battistelle RAG (2017) Environmental analysis of organic waste treatment focusing on composting scenarios. J Clean Prod 155:229–237
63. Al-Madbouh S, Al-Khatib IA, Salahat JI, Jararaa BYA, Ribbe L (2019) Socioeconomic, agricultural, and individual factors influencing farmers' perceptions and willingness of compost production and use: an evidence from Wadi al-Far'a Watershed-Palestine. Environ Monit Assess 191:209. https://doi.org/10.1007/s10661-019-7350-2
64. Al-Sari MI, Sarhan MAA, Al-Khatib IA (2018) Assessment of compost quality and usage for agricultural use: a case study of Hebron. Palestine Environ Monit Assess 190:223. https://doi.org/10.1007/s10661-018-6610-x
65. The World Bank (2019) Economic monitoring report to the ad hoc liaison committee. Washington, D.C. http://documents.worldbank.org/curated/en/410061568815090051/Economic-Monitoring-Report-to-the-Ad-Hoc-Liaison-Committee. Accessed on 7 Jan 2022.
66. UNCTAD (United Nations Conference on Trade and Development) (2015) The Besieged Palestinian Agricultural Sector. New York and Geneva. https://unctad.org/system/files/official-document/gdsapp2015d1_en.pdf. Accessed on 10 Jan 2022.
67. Huber-Humer M, Tintner J, Böhm K, Lechner P (2011) Scrutinizing compost properties and their impact on methane oxidation efficiency. Waste Manage 31:871–883. https://doi.org/10.1016/j.wasman.2010.09.023
68. Sadasivam BY, Reddy KR (2014) Landfill methane oxidation in soil and bio-based cover systems: a review. Rev Environ Sci Biotechnol 13(1):79–107. https://doi.org/10.1007/s11157-013-9325-z
69. JK Chetri, RK Rai, KR Reddy (2019) Effect of temperature on CO_2 sequestration by BOF slag in landfill cover. International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE). American Conference on Soil Mechanics and Geotechnical Engineering (XVI PCSMGE), 2502–2509. https://doi.org/10.3233/STAL190320
70. Tanthachoon N, Chiemchaisr C, Chiemchaisr W (2007) Utilization of Municipal Solid Waste Compost as landfill cover soil for reducing greenhouse gas emission. Int J Environ Technol Manag 7(3/4):286–297
71. Hermawansyah Y, Hadinata F, Dewi R, Hanafiah, (2018) A study of compost use as an alternative daily cover in Sukawinatan Landfill Palembang. Int J Geomate 15(51):47–52
72. The Joint Service Council for Solid Waste Management for Hebron and Bethlehem (JSC-H&B) (2021) Supply of cover materials to cover the waste at Al-Menya landfill. Tender documents no. 10/2021. Unpublished document. Palestine.
73. Ali AM, Nawaz AM, Al-Turaif HA, Shahzad K (2020) The economic and environmental analysis of energy production from slaughterhouse waste in Saudi Arabia. Environ Dev Sustain 23:1–18. https://doi.org/10.1007/s10668-020-00770-6
74. Shah GM, Tufail N, Bakhat HF, Imran M, Murtaza B, Farooq ABU, Saeed F, Waqar A, Rashid MI (2017) Anaerobic degradation of municipal organic waste among others composting techniques improves N cycling through waste-soil-plant continuum. J Soil Sci Plant Nutr 17(2):529–542. https://doi.org/10.4067/S0718-95162017005000038
75. Rashid MI, Shahzad k, (2021) Food waste recycling for compost production and its economic and environmental assessment as circular economy indicators of solid waste management. J Clean Prod 317:128467. https://doi.org/10.1016/j.jclepro.2021.128467
76. Van Iperen International B.V (2021). Iperen Humic 12 + 3 Liquid. https://www.vaniperen.com/products/iperen-humic-12-3-liquid/. Accessed on 20 Nov 2021.
77. Eurosolids Nederland (2020a) Ammonium Sulphate Horticultural Grade data sheet. The Netherland. https://www.eurosolids.com/products/ammonium-sulphate/. Accessed on 20 Nov 2021.
78. TAS (2021) Triple super phosphate (TSP). https://www.tasflowrance.com/product/triple-super-phosphate-tsp#. Accessed on 20 Nov 2021.
79. Eurosolids Netherland (2020b) The Netherland. https://www.eurosolids.com/products/potassium-sulphate-sop/. Accessed on 20 Oct 2021.
80. SBTAC (Sharabati Brothers Trading and Agricultural Company) (2021) Prices of chemical fertilizers.
81. The Joint Service Council for Solid Waste Management for Hebron and Bethlehem (JSC-H&B) (2011) Southern West Bank solid waste management project (SWBSWMP), contract agreement for construction of sanitary landfill at Al-Menya, southern West Bank. Unpublished document. Palestine
82. ALPC (Al-Menya Landfill Private Company) (2016) Landfill survey report, survey no. 6. Unpublished report. Bethlehem, Palestine.
83. Shanker R, Khan D, Hossain R et al (2022) Plastic waste recycling: existing Indian scenario and future opportunities. Int J Environ Sci Technol. https://doi.org/10.1007/s13762-022-04079-x
84. Elshaboury N, Al-Sakkaf A, Abdelkader ME, Alfalah G (2022) Construction and demolition waste management research: a science mapping analysis. Int J Environ Res Public Health 19:4496. https://doi.org/10.3390/ijerph19084496
85. Mohite AS, Jagtap AR, Avhad MS, More AP (2022) Recycling of major agriculture crop residues and its application in polymer industry: a review in the context of waste to energy nexus. Energy Nexus 7:100134. https://doi.org/10.1016/j.nexus.2022.100134
86. Allievi L, Marchesini A, Salardi C, Piano V, Ferrari A (1993) Plant quality and soil residual fertility six years after a compost treatment. Bioresour Technol 43(1):85–89. https://doi.org/10.1016/0960-8524(93)90088-S
87. Jaza Foleface AJ (2009) The substitution of mineral fertilizers by compost from household waste in Cameroon: economic analysis with a partial equilibrium model. Waste Manag Res 27(3):207–223. https://doi.org/10.1177/0734242X08090403
88. Ddiba D, Andersson K, Rosemarin A, Schulte-Herbrüggen H, Dickin S (2022) The circular economy potential of urban organic waste streams in low- and middle-income countries. Environ Dev Sustain 24:1116–1144. https://doi.org/10.1007/s10668-021-01487-w
89. Sardarmehni M, Levis JW, Barlaz MA (2021) What is the best end use for compost derived from the organic fraction of municipal solid waste? Environ Sci Technol 55:73–81. https://doi.org/10.1021/acs.est.0c04997

90. The Palestinian Central Bureau of Statistics (PCBS), and the Palestinian Water Authority (PWA) (2022) Press Release on the Occasion of World Water Day, Ramallah, Palestine.
91. Lotter D (2015) Facing food insecurity in Africa: why, after 30 years of work in organic agriculture, I am promoting the use of synthetic fertilizers and herbicides in small scale staple crop production. Agric Hum Values 32(1):111–118
92. Reganold JP, Wachter JM (2016) Organic agriculture in the twenty-first century. Nature Plants 2(2):1–8
93. Hui L, Wen-ting F, Xin-hua H, Ping Z, Hong-jun G, Nan S, Minggang X (2017) Chemical fertilizers could be completely replaced by manure to maintain high maize yield and soil organic carbon (SOC) when SOC reaches a threshold in the Northeast China Plain. J Integr Agric 16(4):937–946
94. Iqbal A, Ali I, Yuan P, Khan R, Liang H, Wei S, Jiang L (2022) Combined application of manure and chemical fertilizers alters soil environmental variables and improves soil fungal community composition and rice grain yield. Front Microbiol 13:856355. https://doi.org/10.3389/fmicb.2022.856355
95. Timsina J (2018) Can organic sources of nutrients increase crop yields to meet global food demand? Agronomy 8(10):214
96. Heydarzadeh S, Arena C, Vitale E, Rahimi A, Mirzapour M, Nasar J, Kisaka O, Sow S, Ranjan S, Gitari H (2023) Impact of different fertilizer sources under supplemental irrigation and rainfed conditions on eco-physiological responses and yield characteristics of dragon's head (Lallemantia iberica). Plants 12:1693. https://doi.org/10.3390/plants12081693
97. The World Bank (2022) World Bank Commodities Price Data (The Pink Sheet). https://thedocs.worldbank.org/en/doc/5d903e848db1d1b83e0ec8f744e55570-0350012021/related/CMO-Pink-Sheet-September-2022.pdf. Accessed 8 July 2023.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Check for updates

RESEARCH ARTICLE

annalen der physik
www.ann-phys.org

Observational Constraints on the Parameters of Hořava–Lifshitz Gravity

*Himanshu Chaudhary, Ujjal Debnath, Shibesh Kumar Jas Pacif, Niyaz Uddin Molla, Ghulam Mustafa, and Sunil Kumar Maurya**

This study investigates the accelerated cosmic expansion within the Hořava–Lifshitz (HL) Model. To constrain the cosmological parameters of this model, 17 Baryon Acoustic Oscillation points, 31 Cosmic Chronometer points, 40 Type Ia Supernovae points, 24 quasar Hubble diagram points, and 162 Gamma Ray Bursts points, along with the latest Hubble constant measurement (R22) are incorporated. r_d is treated as a free parameter to extract H_0 and r_d using late-time datasets, aiming for optimal fitting values in each model. Treating r_d as free improves precision, reduces bias, and enhances dataset compatibility. The obtained values of H_0 and r_d are compared to the ΛCDM model, showing consistency with previous estimates from Planck and SDSS studies. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) favor the Hořava–Lifshitz model, with the ΛCDM model having the lowest AIC. Additionally, ΔAIC and ΔBIC analyses are conducted to assess model preference. Validation using the reduced χ^2_{red} statistic indicates satisfactory fits for the Hořava–Lifshitz model, while recognizing ΛCDM as the preferred model. Extensions of the analysis warrant further investigation.

1. Introduction

Einstein's General theory of relativity (GR) remains the current gold standard for gravitational theories, consistently passing all tests to date.[1–3] Nevertheless, it is important to develop new theories to further test GR and to explore the potential for a theory of gravity beyond GR. Current research interests include finding a valid theory of quantum gravity and explaining cosmological phenomena such as dark energy and dark matter.[4–6] One challenge with GR is that it is not power-counting renormalizable. To address this, Hořava proposed a theory of gravity beyond GR that is both renormalizable and ultraviolet complete.[7] This theory, known as Hořava–Lifshitz (HL) gravity, breaks Lorentz invariance in the ultraviolet (UV) regime by introducing a Lifshitz-type anisotropic scaling between space and time.

In Einstein's gravitational theory, a prominent challenge arises from UV divergence, known as the UV completion problem. Stelle (1976) addressed this by proposing the incorporation of higher-order curvature invariants into gravitational theories.[8] However, this approach encountered difficulties due to the appearance of ghost degrees of freedom associated with higher-order derivatives. Horava (2009) effectively tackled this issue by introducing additional terms featuring higher-order spatial curvature components into the action, leading to what is now termed deformed HL gravity.[9] Under high-energy conditions, HL gravity is formulated by abandoning Lorentz symmetry through a Lifshitz-type rescaling process.[10] This rescaling involves altering the scaling properties of space and time coordinates, characterized by an exponent denoted as z, and d representing the space-time dimension. Lorentz symmetry is broken when $z \neq 1$ and restored only when $z = 1$. Modified HL gravity has garnered significant attention in scientific literature.[11–17] Researchers have explored various aspects, including cosmological implications, such as the Friedmann equation within the deformed HL gravity framework for the FLRW Universe,[18] novel approaches to dark energy like holographic dark energy (HDE),[19] and investigations into black hole properties such as quasi-normal modes.[20] Additionally, studies have delved into models involving ghost dark energy within the context of deformed HL gravity.[21] The HL gravity model has been extensively examined by numerous

H. Chaudhary
Department of Applied Mathematics
Delhi Technological University
Delhi 110042, India

H. Chaudhary, S. K. J. Pacif
Pacif Institute of Cosmology and Selfology (PICS)
Sagara, Sambalpur, Odisha 768224, India

U. Debnath, N. U. Molla
Department of Mathematics
Indian Institute of Engineering Science and Technology
Shibpur, Howrah 711 103, India

G. Mustafa
Department of Physics
Zhejiang Normal University
Jinhua 321004, P. R. China

G. Mustafa
New Uzbekistan University
Mustaqillik ave. 54, Tashkent 100007, Uzbekistan

S. K. Maurya
Department of Mathematical and Physical Sciences
College of Arts and Sciences
University of Nizwa
Nizwa 616, Sultanate of Oman
E-mail: sunil@unizwa.edu.om

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/andp.202400181

DOI: 10.1002/andp.202400181

 © 2024 Wiley-VCH GmbH

researchers,[22–28] exploring its diverse implications and applications across cosmology and gravitational physics.

HL gravity represents a significant endeavor in theoretical physics, aiming to address the quantum gravity problem and rectify the lack of renormalizability inherent in standard general relativity. By providing a UV-complete theory, it seeks to offer a framework capable of describing gravitational interactions at both small quantum scales and large cosmological scales. Despite its ambitious goals, the theory has not been without challenges and criticisms. One major point of contention revolves around its viability as a quantum theory. Critics have raised concerns about the presence of ghost instabilities, which can lead to inconsistencies and unpredictability within the theory's predictions. These challenges have spurred ongoing debate within the scientific community regarding the feasibility and robustness of HL gravity as a fundamental theory of gravity. Nevertheless, researchers remain engaged in studying and refining the HL gravity framework, particularly in its implications for cosmology and black hole physics. Understanding the cosmological consequences of this theory is of particular interest, as it may offer insights into the nature of the universe at large scales. Investigations into black hole properties within this framework also hold promise for elucidating fundamental aspects of gravitational physics. It's important to recognize that the scientific community's understanding of HL gravity is dynamic and subject to change as further research and experimental data become available. Jawad et al.,[29] investigated the dynamical stability and cosmographic characteristics of some cosmic models under deformed HL gravity, matching their results with observational data. They studied cosmographic parameters and thermodynamics with regard to Einstein's gravity and deformed HL gravity associated with Kaniadakis HDE.[30] Additionally, they explored the cosmological consequences of Sharma–Mittal holographic dark energy (SMHDE), analyzing cosmological parameters and thermodynamics.[31] Further, constraining a model with observations is crucial for ensuring its physical relevance and predictive power. Observational constraints help in refining theoretical models by aligning them with empirical data, thus validating or refuting theoretical predictions. Following the theoretical formulation of HL gravity, various datasets were used by ref. [32] to constrain the corresponding cosmological scenarios. Additionally,[33] focused on the running parameter λ of HL gravity, which governs the flow between the UV and the Infra-Red. They found that λ is restricted to $|\lambda - 1| \lesssim 0.02$, with the best-fit value being $|\lambda_{\text{b.f.}} - 1| \approx 0.002$. This observational analysis confines the running parameter λ very close to its IR value of 1. It is important to note that basic versions of HL gravity may encounter perturbative instabilities, as highlighted in previous studies.[34,35] These instabilities pose challenges to the theory's robustness and reliability. Recognizing and addressing these issues is essential for advancing HL gravity and ensuring its consistency with empirical observations and theoretical expectations.

Motivated by the preceding discourse, this study explores into the realm of HL gravity, which emerges as a compelling alternative to Einstein's gravitational theory for several reasons. Both theories aim to describe the behavior of gravity, but they approach the challenge from different perspectives and with different goals in mind. One key motivation for considering HL gravity over Einstein gravity lies in addressing the UV completion problem, a significant challenge in theoretical physics. HL gravity offers a promising avenue for resolving UV divergences by introducing higher-order spatial curvature components into the action, thereby providing a UV-complete theory capable of describing gravitational interactions at both small quantum scales and large cosmological scales. Our primary objective is to meticulously evaluate the compatibility of HL model with observational data, thus deriving rigorous constraints on crucial cosmic parameters.

Harnessing recent measurements of the Hubble parameter, $H(z)$, extracted from cosmic chronometers, alongside a diverse array of datasets spanning Type Ia Supernovae, Gamma-Ray Bursts (GRB), Quasars, and uncorrelated Baryon Acoustic Oscillations (BAO), our analysis offers a comprehensive survey of the cosmic terrain. Additionally, we aim to probe the ongoing cosmological tensions such as H_0 and the sensitivity of r_d. The paper is structured as follows: In Section 2, we present the basic equations of HL Gravity. In Section 3, we focus on the methodology employed to constrain the crucial parameters of HL model. Our study's outcomes are detailed in Section 4, where we present the results obtained. Finally, in Section 5, we conclude the paper, offering discussions on the implications and significance of our findings.

2. Hořava–Lifshitz Gravity and Basic Equations

We assume the Arnowitt–Deser–Misner decomposition, which is provided as in the metric form:

$$ds^2 = g_{ij}\left(dx^i + N^i dt\right)\left(dx^j + N^j dt\right) - N^2 dt^2 \tag{1}$$

In the above equation, N is representing the lapse function, N_i is providing the shift vector, g_{ij} is denoting the spatial metric. For the present analysis, the scaling transformation in the framework of coordinates reads as $t \to l^3 t$ and $x^i \to l x^i$. The action of the HL gravity has two major components, namely, the potential term and the kinetic term. The HL gravity action within the scope of the potential term and the kinetic term is defined as:

$$S_g = S_v + S_k = \int dt d^3x \sqrt{g} N\left(L_v + L_k\right) \tag{2}$$

where, the kinetic term is defined as

$$S_k = \int dt d^3x \sqrt{g} N\left[\frac{2\left(K_{ij}K^{ij} - \lambda K^2\right)}{\kappa^2}\right] \tag{3}$$

where, The extrinsic curvature is provided as

$$K_{ij} = \frac{\dot{g}_{ij} - \Delta_i N_j - \Delta_j N_i}{2N} \tag{4}$$

The number of invariants, while working with the Lagrangian, L_v, can be reduced due to its symmetric property. This symmetry is referred to as detailed balance. By using this detailed balance, the action can be revised as

© 2024 Wiley-VCH GmbH

15213889, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/andp.202400181 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$S_g = \int dtd^3x\sqrt{g}N\left[\frac{2\left(K_{ij}K^{ij}-\lambda K^2\right)}{\kappa^2}+\frac{\kappa^2 C_{ij}C^{ij}}{2\omega^4}-\frac{\kappa^2\mu\epsilon^{ijk}R_{i,j}\Delta_j R^l_k}{2\omega^2\sqrt{g}}+\frac{\kappa^2\mu^2 R_{ij}R^{ij}}{8}-\frac{\kappa^2\mu^2}{8(3\lambda-1)}\left\{\frac{(1-4\lambda)R^2}{4}+\Lambda R-3\Lambda^2\right\}\right] \quad (5)$$

$$\text{where} \quad C^{ij}=\frac{\epsilon^{ijk}\Delta_k\left(R^j_i-\frac{R}{4}\delta^j_i\right)}{\sqrt{g}} \quad (6)$$

is known as the Cotton tensor. Here the covariant derivatives can be determined by the spatial metric $g_{ij}\epsilon^{ijk}$, which is the antisymmetric unit tensor. Here, the parameter λ is the dimensionless constant, and the other parameters κ, ω, and μ are also the constants.

Horava obtained a gravitational action by considering the temporal dependency of the lapse function $N(t)$. For FRW metric with the assumptions $N=1$, $g_{ij}=a^2(t)\gamma_{ij}$, $N^i=0$ and

$$\gamma_{ij}dx^i dx^j=\frac{dr^2}{1-kr^2}+r^2d\Omega_2^2 \quad (7)$$

and by taking the variations of N and g_{ij}, we can obtain the two Field equations which are given below:

$$H^2=\frac{\kappa^2\rho_m}{6(3\lambda-1)}+\frac{\kappa^2}{6(3\lambda-1)}\left[\frac{3\kappa^2\mu^2k^2}{8(3\lambda-1)a^4}+\frac{3\kappa^2\mu^2\Lambda^2}{8(3\lambda-1)}\right]-\frac{\kappa^4\mu^2\Lambda k}{8(3\lambda-1)^2a^2} \quad (8)$$

$$\dot{H}+\frac{3H^2}{2}=-\frac{\kappa^2p_m}{4(3\lambda-1)}-\frac{\kappa^2}{4(3\lambda-1)}\left[\frac{3\kappa^2\mu^2k^2}{8(3\lambda-1)a^4}+\frac{3\kappa^2\mu^2\Lambda^2}{8(3\lambda-1)}\right]-\frac{\kappa^4\mu^2\Lambda k}{8(3\lambda-1)^2a^2} \quad (9)$$

where $k=0,\pm1$ represent the flat, closed, and open universe respectively. In the field equations mentioned above, the term proportional to $\frac{1}{a^4}$ is treated as the "Dark radiation" term and Λ denotes the cosmological constant. Also, $H=\frac{\dot{a}}{a}$ represents the Hubble parameter. Now we assume that the universe is filled with dark matter and we have not taken any external dark energy because the modified gravity can produce dark energy. So the conservation equation of the dark matter (DM) is given by

$$\dot{\rho}_m+3H(\rho_m+p_m)=0 \quad (10)$$

Since the dark matter has negligible pressure (i.e., $p_m\sim0$), so the Equation (10) yields $\rho_m=\rho_{m0}a^{-3}$. Using the dimensionless parameters $\Omega_{m0}\equiv\frac{\rho_{m0}}{3H_0^2}$, $\Omega_{k0}=-\frac{k}{H_0^2}$, $\Omega_{\Lambda0}=\frac{\Lambda}{2H_0^2}$, we obtain (choosing $\kappa^2=1$)

$$H^2(z)=\frac{2H_0^2\Omega_{m0}(1+z)^3}{3\lambda-1}+\frac{\mu^2H_0^2}{16(3\lambda-1)^2}\left[\Omega_{k0}(1+z)^2+3\Omega_{\Lambda0}\right]^2 \quad (11)$$

3. Data and Methodology

In the investigation, a specific set of recent Baryon Acoustic Oscillation (BAO) measurements extracted from multiple galaxy surveys was carefully picked, with a primary focus on data from the Sloan Digital Sky Survey (SDSS).[36–41] To enrich the dataset, it was augmented with valuable contributions from other surveys such as the Dark Energy Survey (DES),[42] the Dark Energy Camera Legacy Survey (DECaLS),[43] and 6dFGS BAO.[44] Recognizing the potential for interdependencies among the selected data points, this concern was diligently addressed. A subset was carefully curated to mitigate highly correlated points, fully appreciating the importance of managing possible correlations within the chosen dataset. To ensure a comprehensive assessment of systematic errors, mock datasets generated from N-body simulations were employed to precisely calculate covariance matrices. Obtaining accurate covariance matrices among measurements derived from diverse observational surveys presented a significant challenge. To address a significant challenge, covariance analysis was utilized as outlined in ref. [45]. Specifically, an approach was adopted where the diagonal elements of the covariance matrix, denoted as C_{ii}, was set equal to the square of the 1σ errors, σ_i^2. In order to introduce correlations among the selected subset of data, the covariance matrix was augmented by incorporating non-diagonal elements while preserving its symmetry. This was achieved by randomly pairing data points, resulting in a total of 12 pairs. For each pair, the non-diagonal elements were determined using the expression $C_{ij}=0.5\sigma_i\sigma_j$, where σ_i and σ_j represent the 1σ errors associated with data points i and j, respectively. Through this methodology, correlations within 56% of the BAO dataset under consideration were effectively stimulated. This approach enabled to capture the interdependence among data points, providing a more comprehensive representation of the underlying structure within the dataset. By incorporating these correlations, the aim was to enhance the accuracy and reliability of the analyses, thereby improving the understanding of the phenomenon under investigation. To constrain the cosmological model parameters, the dataset was expanded by including various additional measurements. First, 30 independent Hubble parameter measurements derived from cosmic chronometers (CC) detailed in refs. [46–49] were incorporated. Additionally, the study integrated recent Pantheon Type Ia Supernovae samples,[50] 24 binned quasar[51] distance modulus data, 162 Gamma-Ray Bursts (GRBs),[52] and the latest Hubble constant measurement (R22) as an additional prior.[53] To analyze these datasets, a nested sampling approach using the Polychord package was adopted,[54] enabling thorough exploration of parameter space for determining optimal fit values. Additionally, the GetDist package was utilized [55] to present the findings clearly and informatively, enhancing the interpretation and visualization of the analysis outcomes. This comprehensive methodology ensured a robust understanding of cosmological parameters while providing insights into the Universe's fundamental properties.

4. Results

Figure 1 shows the 68% and 95% confidence levels for important cosmological parameters in the HL Model. The best-fit values for these parameters are listed in **Table 1**. When we incorporate the

© 2024 Wiley-VCH GmbH

15213889, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/andp.202400181 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

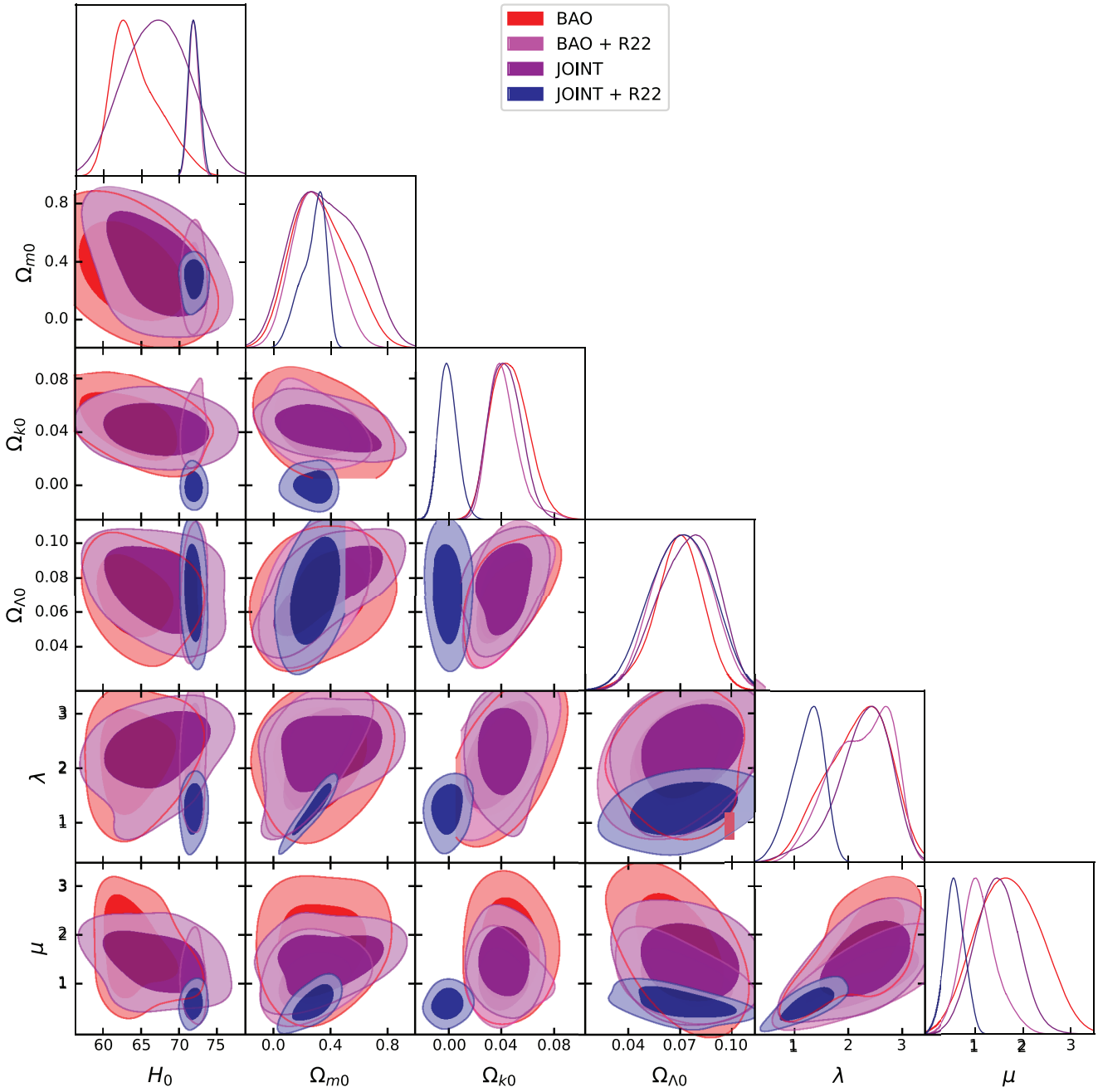


Figure 1. The figure illustrates the posterior distribution of various observational data measurements using the HL Model, highlighting the 1σ and 2σ regions.

R22 prior into the Joint dataset, the resulting value for H_0 deviates from the findings in ref. [56], but closely aligns with the SNIe sample in ref. [53]. Conversely, without the R22 priors and using the Joint dataset alone, the estimated value of H_0 aligns more closely with ref. [56]. These findings indicate that the inclusion of the R22 prior diverges from ref. [56] and aligns with ref. [53] showing the impact of prior choice on cosmological parameter estimation. The determined values for the matter density parameter (Ω_{m0}) and (Ω_{d0}) seem lower compared to those documented in ref. [56]. However, this observation aligns with findings from alternative studies.[57,58] One can observe that the running parameter λ, according to the Joint with R22 analysis, is constrained close to the expected value of 1. Now turning our attention to the Baryon Acoustic Oscillations (BAO) scale, which is a fundamental scale in cosmology derived from the cosmic microwave background, this scale originates from a significant event known as the drag epoch (z_d), marking a pivotal moment when baryons and photons transitioned into distinct entities. The BAO scale (r_d) is essentially a measure of the cosmic sound horizon at z_d, which reflects the separation distance between these baryons and

© 2024 Wiley-VCH GmbH

15213889, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/andp.202400181 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1. Summary of the MCMC results.

MCMC Results						
Model	Parameter	Priors	BAO	BAO + R22	Joint	Joint + R22
ΛCDM Model	H_0	[50.,80.]	$70.256500^{\pm 9.206346}_{\pm 5.502574}$	$73.798696^{\pm 2.543240}_{\pm 1.315504}$	$69.837362^{\pm 2.478047}_{\pm 1.204873}$	$71.508817^{\pm 1.614354}_{\pm 0.882681}$
	Ω_{m0}	[0.,1.]	$0.270161^{\pm 0.033280}_{\pm 0.013827}$	$0.268255^{\pm 0.043787}_{\pm 0.016790}$	$0.274526^{\pm 0.021367}_{\pm 0.009858}$	$0.270609^{\pm 0.022243}_{\pm 0.009678}$
	r_d(Mpc)	[100.,200.]	$145.807497^{\pm 13.994434}_{\pm 9.060325}$	$138.345762^{\pm 4.939114}_{\pm 2.451777}$	$145.811932^{\pm 4.545176}_{\pm 2.347736}$	$142.591327^{\pm 3.951717}_{\pm 1.850175}$
	r_d/r_{fid}	[0.9,1.1]	$0.970661^{\pm 0.068973}_{\pm 0.055989}$	$0.930500^{\pm 0.028238}_{\pm 0.021272}$	$0.974064^{\pm 0.057975}_{\pm 0.033510}$	$0.947662^{\pm 0.043021}_{\pm 0.028335}$
	H_0	[50.,80.]	$65.700835^{\pm 8.188232}_{\pm 6.213905}$	$73.764424^{\pm 2.238343}_{\pm 1.206257}$	$69.083097^{\pm 13.449031}_{\pm 9.523705}$	$73.841896^{\pm 2.325034}_{\pm 1.470940}$
	Ω_{m0}	[0.,1.]	$0.307520^{\pm 0.374679}_{\pm 0.246232}$	$0.313934^{\pm 0.279260}_{\pm 0.193706}$	$0.301674^{\pm 0.441591}_{\pm 0.290568}$	$0.316901^{\pm 0.251704}_{\pm 0.121592}$
	Ω_{k0}	[0.,0.1]	$0.045949^{\pm 0.019647}_{\pm 0.015391}$	$0.041633^{\pm 0.015582}_{\pm 0.011413}$	$0.042779^{\pm 0.017139}_{\pm 0.013332}$	$-0.000867^{\pm 0.011261}_{\pm 0.006873}$
Hořava–Lifshitz Model	$\Omega_{\Lambda 0}$	[0.,0.1]	$0.069321^{\pm 0.031470}_{\pm 0.011309}$	$0.071208^{\pm 0.025913}_{\pm 0.016735}$	$0.074617^{\pm 0.032586}_{\pm 0.017924}$	$0.070843^{\pm 0.032504}_{\pm 0.018968}$
	λ	[0.,1.]	$2.179575^{\pm 1.092257}_{\pm 0.718167}$	$2.250670^{\pm 1.174342}_{\pm 0.613784}$	$2.286921^{\pm 1.402826}_{\pm 0.450757}$	$1.249560^{\pm 0.687458}_{\pm 0.309072}$
	μ	[0.,3.]	$1.709372^{\pm 1.036271}_{\pm 0.696013}$	$1.090299^{\pm 0.706884}_{\pm 0.313413}$	$1.457249^{\pm 0.864455}_{\pm 0.445976}$	$0.581412^{\pm 0.432033}_{\pm 0.201536}$
	r_d (Mpc)	[100.,200.]	$148.489236^{\pm 14.876427}_{\pm 8.942271}$	$147.612337^{\pm 13.663930}_{\pm 9.931756}$	$146.021893^{\pm 5.394418}_{\pm 2.757663}$	$146.862925^{\pm 5.019343}_{\pm 3.047961}$
	r_d/r_{fid}	[0.9,1.1]	$1.004965^{\pm 0.099047}_{\pm 0.058313}$	$0.993782^{\pm 0.086915}_{\pm 0.064095}$	$0.979637^{\pm 0.032327}_{\pm 0.020247}$	$0.991474^{\pm 0.034560}_{\pm 0.020670}$

photons. To calculate r_d, we consider the integral of the ratio between the speed of sound (c_s) and the Hubble parameter (H) across a range of redshifts, starting from z_d and extending indefinitely. The speed of sound, c_s, is determined by the fluctuations in pressure (δp_γ) and density ($\delta\rho_B$ and $\delta\rho_\gamma$) within the photon and baryon components. Mathematically, this simplifies to $\frac{1}{\sqrt{3(1+R)}}$, where R represents the ratio of baryon density fluctuations to photon density fluctuations ($R \equiv \frac{\delta\rho_B}{\delta\rho_\gamma} = \frac{3\rho_B}{4\rho_\gamma}$). Observational data, as documented in ref. [56], has provided us with the precise redshift of the drag epoch, $z_d = 1059.94 \pm 0.30$. In the context of a ΛCDM model, the BAO scale r_d has been estimated to be 147.09 ± 0.26 megaparsecs (Mpc) based on measurements by ref. [56]. **Figure 2** shows the posterior distribution for the $r_d - H_0$ contour plane of the HL Model. In the HL model, the BAO scale derived solely from BAO datasets is 148.814 ± 10.558 Mpc. Upon including the R22 prior exclusively into the BAO dataset, the sound horizon at the drag epoch becomes 138.907 ± 2.129 Mpc. For the Joint dataset, the BAO scale is 145.843 ± 2.47 Mpc. Integrating the R22 prior into the full dataset yields $r_d = 142.742 \pm 1.658$ Mpc. The findings indicate that when incorporating the joint dataset, the obtained value of r_d in each dark energy model closely agrees with the values obtained in ref. [56]. However, when combining the R22 priors with the joint dataset, the obtained value of r_d aligns closely with refs. [59, 60] introduces an insightful perspective by employing binning and Gaussian methods to amalgamate measurements of 2D BAO and SNIa data. It's important to recognize that the outcomes we derive are closely tied to the initial range of priors we make about the BAO scale (r_d) and the Hubble constant (H_0). These priors significantly impact the values we estimate, particularly when considering the $r_d - H_0$ contour plot. An interesting observation emerges when we omit the R22 prior from our analysis: in such cases, the results for H_0 and r_d tend to better align with findings from the Planck and SDSS experiments.

When assessing various cosmological models, we employ both the Akaike Information Criterion (AIC)[57] and the Bayesian Information Criterion (BIC) for evaluation purposes. The AIC and BIC are statistical measures used to assess the goodness of fit of a model to a given dataset while penalizing for the number of parameters in the model. The AIC is calculated using the formula: $\mathrm{AIC} = -2\ln\left(\mathcal{L}_{\max}\right) + 2k + \frac{2k(2k+1)}{N_{\mathrm{tot}}-k-1}$. Here, $\mathcal{L}_{\max}$ represents the maximum likelihood of the data, N_{tot} is the total number of data points, and k is the number of parameters in the model. When the total number of data points (N_{tot}) is large, the expression simplifies to: $\mathrm{AIC} \simeq -2\ln\left(\mathcal{L}_{\max}\right) + 2k$ which is the conventional form of the AIC criterion. This criterion provides a trade-off between the goodness of fit and the complexity of the model. On the other hand, the BIC is defined as: $\mathrm{BIC} = -2\ln\left(\mathcal{L}_{\max}\right) + k\ln N_{\mathrm{tot}}$. The BIC also evaluates the goodness of fit while penalizing for the number of parameters, but it imposes a stronger penalty for additional parameters compared to the AIC, particularly for smaller sample sizes. To compare different models, we use the differences in AIC and BIC values, denoted as ΔAIC and ΔBIC, respectively: $\Delta\mathrm{AIC} = \mathrm{AIC}_{\mathrm{HL\ Model}} - \mathrm{AIC}_{\Lambda\mathrm{CDM\ Model}}$ and $\Delta\mathrm{BIC} = \mathrm{BIC}_{\mathrm{HL\ Model}} - \mathrm{BIC}_{\Lambda\mathrm{CDM\ Model}}$ These differences help to identify the relative performance of models, with lower values indicating a better fit relative to the reference model. The interpretation of ΔAIC and ΔBIC values is as follows: $-\Delta\mathrm{AIC}$ or $\Delta\mathrm{BIC} \leq 2$: Substantial evidence for the model. $-\Delta\mathrm{AIC}$ or ΔBIC between 4 and 7: Considerable evidence against the model. $-\Delta\mathrm{AIC}$ or $\Delta\mathrm{BIC} \geq 10$: Strong evidence against the model. When comparing the HL Model with the standard ΛCDM model, it's important to note that both proposed extensions encompass ΛCDM as a subset, differing only by 4 degrees of freedom. This aspect allows for the application of conventional statistical tests. Our evaluation primarily employs the reduced chi-square statistic, denoted as $\chi^2_{\mathrm{red}} = \chi^2/\mathrm{Dof}$, where "Dof" represents the degrees of freedom of the model and χ^2

© 2024 Wiley-VCH GmbH

15213889, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/andp.202400181 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

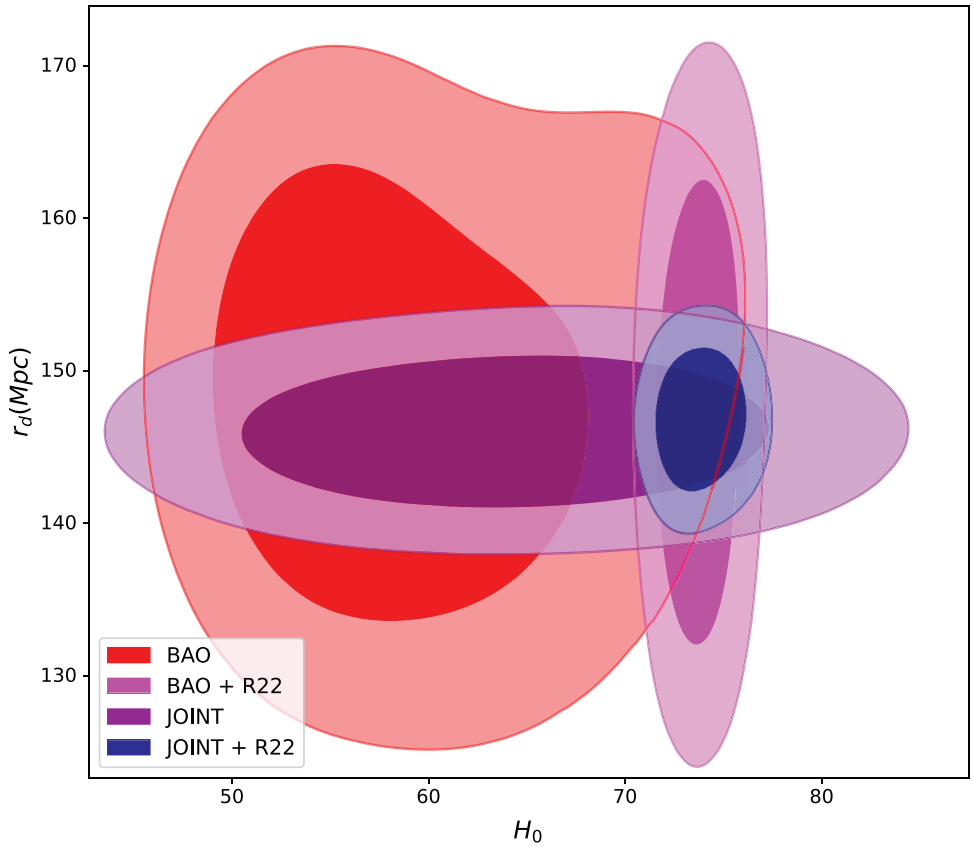


Figure 2. The figure illustrates the posterior distribution of various observational data measurements using the HL Model, highlighting the 1σ and 2σ regions.

signifies the weighted sum of squared deviations from the expected
values.

Based on the provided in **Table 2**, we can conduct a comparative study between the ΛCDM Model and HL Model using the metrics of reduced chi-square (χ^2_{red}), Akaike Information Criterion (AIC), ΔAIC, Bayesian Information Criterion (BIC), and ΔBIC. The reduced chi-square value is a measure of how well the model fits the data. A value closer to 1 indicates a better fit. Both models have χ^2_{red} values close to 1, suggesting that both models fit the data well. However, the ΛCDM Model has a slightly lower χ^2_{red} value, indicating a marginally better fit. In terms of the AIC, the ΛCDM Model scores 277.38, whereas the HL Model scores slightly higher at 279.13, resulting in a ΔAIC of 1.75. This indicates substantial evidence for the HL Model, though it is less favored compared to the ΛCDM Model. For the BIC, the ΛCDM Model has a score of 277.59, while the HL Model has a higher score of 280.51, leading to a ΔBIC of 2.92. This ΔBIC suggests positive evidence against the HL Model, indicating a stronger preference for the ΛCDM Model.

Table 2. Summary of χ^2_{red}, AIC, ΔAIC, BIC, and ΔBIC.

Model	χ^2_{red}	AIC	ΔAIC	BIC	ΔBIC
ΛCDM Model	0.949	277.38	0	277.59	0
Hořava–Lifshitz Model	0.961	279.13	1.75	280.51	2.92

5. Conclusion

In conclusion, our study focused on investigating the accelerated cosmic expansion in the late Universe within the context of the HL Model. This framework offers an alternative perspective to general relativity by incorporating anisotropic scaling at ultraviolet scales. We analyzed 24 Baryon Acoustic Oscillation (BAO) points, along with 31 uncorrelated data points obtained from Cosmic Chronometers methods. We also incorporated 40 points from Type Ia supernovae, 24 points from the Hubble diagram for quasars, and a substantial dataset consisting of 162 points sourced from Gamma Ray Bursts. Furthermore, our study included the latest measurement of the Hubble constant conducted by R22. Our primary objective was to determine the optimal fitting values for every cosmological parameter within each dark energy model. By treating the sound horizon r_d as a free parameter instead of relying on prior information from the Cosmic Microwave Background (CMB), we gain several benefits. First, it reduces bias by sidestepping potentially restrictive CMB priors, ensuring a more impartial estimation of r_d and other cosmological parameters. Second, it enhances precision by allowing late-time data, such as Baryon Acoustic Oscillations (BAO) or Type Ia Supernovae (SNIe), to directly inform the

© 2024 Wiley-VCH GmbH

15213889, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/andp.202400181 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

measurement of r_d, leading to more accurate results. Moreover, this approach promotes compatibility between datasets from different cosmic epochs, facilitating a more holistic understanding of the Universe's evolution. We determined H_0 and r_d for the HL Model relative to the ΛCDM model as follows: In the ΛCDM model, our analysis yields $H_0 = 69.828145 \pm 1.009964$ km s^{-1} Mpc^{-1} and $r_d = 146.826212 \pm 2.201612$ Mpc. In the HL model, we obtain $H_0 = 69.966349 \pm 1.403864$ km s^{-1}Mpc^{-1} and $r_d = 145.843487 \pm 2.470309$ Mpc. Importantly, our findings underscore that the values of H_0 and r_d based on low-redshift measurements align with early Planck and SDSS estimations. While both the ΛCDM Model and the HL Model fit the data reasonably well, the ΛCDM Model is preferred based on the statistical criteria provided. The ΛCDM Model demonstrates a marginally better fit with a lower reduced chi-square value and is more strongly supported by both the AIC and BIC measures. The ΔAIC indicates substantial evidence for the HL Model, but it is still less favored compared to the ΛCDM Model. Additionally, the ΔBIC provides positive evidence against the HL Model, reinforcing the stronger preference for the ΛCDM Model for the given dataset. Therefore, the ΛCDM Model is favored over the HL Model. These findings underscore the importance of considering alternative cosmological models beyond the standard framework. While the ΛCDM model remains favored due to its simplicity and good fit, the extensions present viable candidates worthy of further investigation. To achieve a better understanding of the fundamental characteristics and development of the Universe, cosmological models must be continually examined and improved.

Acknowledgements

N.U.M. would like to thank CSIR, Govt. of India for providing Senior Research Fellowship (No. 08/003(0141)/2020-EMR-I).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Keywords

ΛCDM model, cosmology, Hořava–Lifshitz gravity

Received: June 13, 2024
Revised: June 20, 2024
Published online:

[1] C. M. Will, *Theory and experiment in gravitational physics*, Cambridge University Press, Cambridge **2018**.
[2] E. Berti, E. Barausse, V. Cardoso, L. Gualtieri, P. Pani, U. Sperhake, L. C. Stein, N. Wex, K. Yagi, T. Baker, C. P. Burgess, F. S. Coelho, D. Doneva, A. De Felice, P. G. Ferreira, P. C. C. Freire, J. Healy, C. Herdeiro, M. Horbatsch, B. Kleihaus, A. Klein, K. Kokkotas, J. Kunz, P. Laguna, R. N. Lang, T. G. F. Li, T. Littenberg, A. Matas, S. Mirshekari, H. Okawa, et al., *Class. Quant. Grav.* **2015**, *32*, 243001.
[3] C. M. Will, *Living Rev. Relativ.* **2014**, *17*, 1.
[4] B. Jain, J. Khoury, *Ann. Phys.* **2010**, *325*, 1479.
[5] T. Clifton, P. G. Ferreira, A. Padilla, C. Skordis, *Phys. Rep.* **2012**, *513*, 1.
[6] K. Koyama, *Rep. Prog. Phys.* **2016**, *79*, 046902.
[7] P. Horava, *Phys. Rev. D* **2009**, *79*, 084008.
[8] K. S. Stelle, *Phys. Rev. D* **1977**, *16*, 953.
[9] P. Hořava, *Phys. Rev. D* **2009**, *79*, 084008.
[10] M. Minamitsuji, *Phys. Lett. B* **2010**, *684*, 194.
[11] J. I. Radkovski, S. M. Sibiryakov, *arXiv preprint arXiv:2306.00102* **2023**.
[12] A. Jawad, A. M. Sultan, S. Rani, *Symmetry* **2023**, *15*, 824.
[13] S. Maity, U. Debnath, *Int. J. Geom. Methods Mod. Phys.* **2020**, *17*, 2050170.
[14] K. Kim, J. J. Oh, C. Park, E. J. Son, *Phys. Rev. D* **2021**, *103*, 044052.
[15] A. N. Tawfik, A. M. Diab, E. Abou El Dahab, *Int. J. Geom. Methods Mod. Phys.* **2017**, *26*, 1750166.
[16] Y. Misonoh, M. Fukushima, S. Miyashita, *Phys. Rev. D* **2017**, *95*, 044044.
[17] J.-W. Lu, Y.-B. Wu, J. Xiao, C.-J. Lu, M.-L. Liu, *Int. J. Mod. Phys. A* **2016**, *31*, 1650110.
[18] S.-W. Wei, Y.-X. Liu, Y.-Q. Wang, *Commun. Theor. Phys.* **2011**, *56*, 455.
[19] A. Jawad, S. Chattopadhyay, *Astrophys. Space Sci.* **2014**, *353*, 293.
[20] S. Chen, J. Jing, *Phys. Lett. B* **2010**, *687*, 124.
[21] A. Sheykhi, S. Ghaffari, H. Moradpour, *Int. J. Mod. Phys. D* **2019**, *28*, 1950080.
[22] M. Chaichian, S. Nojiri, S. D. Odintsov, M. Oksanen, A. Tureanu, *Class. Quant. Grav.* **2010**, *27*, 185021.
[23] S. Carloni, M. Chaichian, S. Nojiri, S. D. Odintsov, M. Oksanen, A. Tureanu, *Phys. Rev. D* **2010**, *82*, 065020.
[24] E. Elizalde, S. Nojiri, S. Odintsov, D. Saez-Gomez, *Eur. Phys. J. C* **2010**, *70*, 351.
[25] S. Nojiri, S. D. Odintsov, *Phys. Rev. D* **2010**, *81*, 043001.
[26] S. Nojiri, S. D. Odintsov, *Phys. Lett. B* **2010**, *691*, 60.
[27] S. Nojiri, S. D. Odintsov, *Phys. Rev. D* **2011**, *83*, 023001.
[28] K. Bamba, S. Capozziello, S. Nojiri, S. D. Odintsov, *Astrophys. Space Sci.* **2012**, *342*, 155.
[29] A. Jawad, M. Usman, *Eur. Phys. J. Plus* **2023**, *138*, 1.
[30] A. Jawad, M. Usman, *Eur. Phys. J. Plus* **2023**, *138*, 1.
[31] V. C. Dubey, U. K. Sharma, A. Pradhan, *Int. J. Geom. Meth. Modern Phys.* **2021**, *18*, 2150002.
[32] S. Dutta, E. N. Saridakis, *J. Cosmol. Astropart. Phys.* **2010**, *2010*, 013.
[33] S. Dutta, E. Saridakis, *J. Cosmol. Astropart. Phys* **2010**, *2010*, 013.
[34] C. Bogdanos, E. N. Saridakis, *Class. Quant. Grav.* **2010**, *27*, 075005.
[35] C. Charmousis, G. Niz, A. Padilla, P. M. Saffin, *J. High Energy Phys.* **2009**, *2009*, 070.
[36] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, M. Manera, *Mon. Not. R. Astron. Soc.* **2015**, *449*, 835.
[37] S. Alam, M. Ata, S. Bailey, F. Beutler, D. Bizyaev, J. A. Blazek, A. S. Bolton, J. R. Brownstein, A. Burden, C.-H. Chuang, J. Comparat, A. J. Cuesta, K. S. Dawson, D. J. Eisenstein, S. Escoffier, H. Gil-Marín, J. N. Grieb, N. Hand, S. Ho, K. Kinemuchi, D. Kirkby, F. Kitaura, E. Malanushenko, V. Malanushenko, C. Maraston, C. K. McBride, R. C. Nichol, M. D. Olmstead, D. Oravetz, N. Padmanabhan, et al., *Mon. Not. R. Astron. Soc.* **2017**, *470*, 2617.
[38] H. Gil-Marín, J. E. Bautista, R. Paviot, M. Vargas-Magaña, S. de La Torre, S. Fromenteau, S. Alam, S. Ávila, E. Burtin, C.-H. Chuang, K. S. Dawson, J. Hou, A. de Mattia, F. G. Mohammad, E.-M. Müller, S. Nadathur, R. Neveux, W. J. Percival, A. Raichoor, M. Rezaie, A. J. Ross, G. Rossi, V. Ruhlmann-Kleider, A. Smith, A. Tamone, J. L. Tinker, R. Tojeiro, Y. Wang, G.-B. Zhao, C. Zhao, et al., *Mon. Not. R. Astron. Soc.* **2020**, *498*, 2492.

© 2024 Wiley-VCH GmbH

15213889, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/andp.202400181 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[39] A. Raichoor, A. De Mattia, A. J. Ross, C. Zhao, S. Alam, S. Avila, J. Bautista, J. Brinkmann, J. R. Brownstein, E. Burtin, M. J. Chapman, C.-H. Chuang, J. Comparat, K. S. Dawson, A. Dey, H. du Mas des Bourboux, J. Elvin-Poole, V. Gonzalez-Perez, C. Gorgoni, J.-P. Kneib, H. Kong, D. Lang, J. Moustakas, A. D. Myers, E.-M. Müller, S. Nadathur, J. A. Newman, W. J. Percival, M. Rezaie, G. Rossi, et al., *Mon. Not. R. Astron. Soc.* **2021**, *500*, 3254.
[40] J. Hou, A. G. Sánchez, A. J. Ross, A. Smith, R. Neveux, J. Bautista, E. Burtin, C. Zhao, R. Scoccimarro, K. S. Dawson, A. de Mattia, A. de la Macorra, H. du Mas des Bourboux, D. J. Eisenstein, H. Gil-Marín, B. W. Lyke, F. G. Mohammad, E.-M. Mueller, W. J. Percival, M. Vargas Magaña, G. Rossi, P. Zarrouk, G.-B. Zhao, J. Brinkmann, J. R. Brownstein, C.-H. Chuang, A. D. Myers, J. A. Newman, D. P. Schneider, M. Vivek, *Mon. Not. R. Astron. Soc.* **2021**, *500*, 1201.
[41] H. D. M. Des Bourboux, J. Rich, A. Font-Ribera, V. de Sainte Agathe, J. Farr, T. Etourneau, J.-M. Le Goff, A. Cuceu, C. Balland, J. E. Bautista, M. Blomqvist, J. Brinkmann, J. R. Brownstein, S. Chabanier, E. Chaussidon, K. Dawson, A. X. González-Morales, J. Guy, B. W. Lyke, A. de la Macorra, E.-M. Mueller, A. D. Myers, C. Nitschelm, A. M. Gutiérrez, N. Palanque-Delabrouille, J. Parker, W. J. Percival, I. Pérez-Ráfols, P. Petitjean, M. M. Pieri, et al., *Astrophys. J.* **2020**, *901*, 153.
[42] T. Abbott, M. Aguena, S. Allam, A. Amon, F. Andrade-Oliveira, J. Asorey, S. Avila, G. Bernstein, E. Bertin, A. Brandao-Souza, D. Brooks, D. L. Burke, J. Calcino, H. Camacho, A. Carnero Rosell, D. Carollo, M. Carrasco Kind, J. Carretero, F. J. Castander, R. Cawthon, K. C. Chan, A. Choi, C. Conselice, M. Costanzi, M. Crocce, L. N. da Costa, M. E. S. Pereira, T. M. Davis, J. De Vicente, S. Desai, et al., *Phys. Rev. D* **2022**, *105*, 043512.
[43] S. Sridhar, Y.-S. Song, A. J. Ross, R. Zhou, J. A. Newman, C.-H. Chuang, R. Blum, E. Gaztanaga, M. Landriau, F. Prada, *Astrophys. J.* **2020**, *904*, 69.
[44] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, F. Watson, *Mon. Not. R. Astron. Soc.* **2011**, *416*, 3017.
[45] L. Kazantzidis, L. Perivolaropoulos, *Phys. Rev. D* **2018**, *97*, 103503.
[46] M. Moresco, *Mon. Not. R. Astron. Soc. Lett.* **2015**, *450*, L16.
[47] M. Moresco, L. Pozzetti, A. Cimatti, R. Jimenez, C. Maraston, L. Verde, D. Thomas, A. Citro, R. Tojeiro, D. Wilkinson, *J. Cosmol. Astropart. Phys.* **2016**, *2016*, 014.
[48] M. Moresco, L. Verde, L. Pozzetti, R. Jimenez, A. Cimatti, *J. Cosmol. Astropart. Phys.* **2012**, *2012*, 053.
[49] M. Moresco, A. Cimatti, R. Jimenez, L. Pozzetti, G. Zamorani, M. Bolzonella, J. Dunlop, F. Lamareille, M. Mignoli, H. Pearce, P. Rosati, D. Stern, L. Verde, E. Zucca, C. M. Carollo, T. Contini, J.-P. Kneib, O. Le Fevre, S. J. Lilly, V. Mainieri, A. Renzini, M. Scodeggio, I. Balestra, R. Gobat, R. McLure, S. Bardelli, A. Bongiorno, K. Caputi, O. Cucciati, S. de la Torre, et al., *J. Cosmol. Astropart. Phys.* **2012**, *2012*, 006.
[50] M. Smith, C. B. D'Andrea, M. Sullivan, A. Möller, R. Nichol, R. Thomas, A. Kim, M. Sako, F. Castander, A. Filippenko, R. J. Foley, L. Galbany, S. González-Gaitán, E. Kasai, R. P. Kirshner, C. Lidman, D. Scolnic, D. Brout, T. M. Davis, R. R. Gupta, S. R. Hinton, R. Kessler, J. Lasker, E. Macaulay, R. C. Wolf, B. Zhang, J. Asorey, A. Avelino, B. A. Bassett, J. Calcino, et al., *Astro. J.* **2020**, *160*, 267.
[51] C. Roberts, K. Horne, A. O. Hodson, A. D. Leggat, *arXiv preprint arXiv:1711.10369* **2017**.
[52] M. Demianski, E. Piedipalumbo, D. Sawant, L. Amati, *Astron. Astrophys.* **2017**, *598*, A113.
[53] A. G. Riess, W. Yuan, L. M. Macri, D. Scolnic, D. Brout, S. Casertano, D. O. Jones, Y. Murakami, G. S. Anand, L. Breuval, T. G. Brink, A. V. Filippenko, S. Hoffmann, S. W. Jha, W. D. Kenworthy, G. Anand, J. Mackenty, B. E. Stahl, W. Zheng, *Astrophys. J. letters* **2022**, *934*, L7.
[54] W. Handley, M. Hobson, A. Lasenby, *Mon. Not. R. Astron. Soc. Lett.* **2015**, *450*, L61.
[55] A. Lewis, *arXiv preprint arXiv:1910.13970* **2019**.
[56] P. Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. Banday, R. Barreiro, N. Bartolo, S. Basak, R. Battye, K. Benabed, J.-P. Bernard, M. Bersanelli, P. Bielewicz, J. J. Bock, J. R. Bond, J. Borrill, F. R. Bouchet, F. Boulanger, M. Bucher, C. Burigana, R. C. Butler, E. Calabrese, J.-F. Cardoso, J. Carron, A. Challinor, H. C. Chiang, J. Chluba, et al., *Astronomy Astrophys.* **2020**.
[57] R. C. Nunes, S. K. Yadav, J. Jesus, A. Bernui, *Mon. Not. R. Astron. Soc.* **2020**, *497*, 2133.
[58] R. C. Nunes, A. Bernui, *Eur. Phys. J. C* **2020**, *80*, 1.
[59] L. Verde, J. L. Bernal, A. F. Heavens, R. Jimenez, *Mon. Not. R. Astron. Soc.* **2017**, *467*, 731.
[60] T. Lemos, Ruchika, J. C. Carvalho, J. Alcaniz, *Eur. Phys. J. C* **2023**, *83*, 495.

© 2024 Wiley-VCH GmbH

15213889, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/andp.202400181 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Check for updates

Received: 4 April 2024 | Revised: 9 June 2024 | Accepted: 18 June 2024

DOI: 10.1002/cta.4161

ORIGINAL ARTICLE

WILEY

Performance analysis of fractional-order modified SRF PLL under grid abnormalities

Oinam Lotika Devi | **Alka Singh**

Electrical Engineering Department, Delhi Technological University, Delhi, India

Correspondence
Alka Singh, Electrical Engineering Department, Delhi Technological University, Delhi, India.
Email: alkasingh.eed@gmail.com

Abstract

This article proposes two different structures of fractional-order modified synchronous reference frame phase-locked loop (MSRF PLL) and discusses their performance under different grid abnormalities. Phase-locked loop (PLL) is a type of closed-loop feedback control system that ensures phase and frequency coherence between its input and output signals. The basic synchronous reference frame phase-locked loop (SRF-PLL) is a conventional synchronization technique that is frequently employed in grid-connected systems for power electronic converters. The SRF-PLL offers rapid and precise phase/frequency detection under ideal grid environments. However, its performance is severely hampered under unbalanced and distorted grid environments. This paper discusses two new configurations of fractional-order (FO) modified SRF (MSRF), one with fractional order only in additional low-pass filter of first order (FO-LP) and another fractional order in both first-order low-pass filter and PI (FO-LPFO-PI) of MSRF. These controllers are assembled using FOs "a" and "b" with limits as $0 < a < 2$ and $0 < b < 2$. The performance analysis of proposed FO MSRFs is done under grid abnormalities like voltage sag and swell, polluted grid supply, frequency change, phase change, and variables for dc offset. The outcomes of simulation are acquired using FO modeling and control (FOMCON) toolbox for MATLAB/SIMULINK, and the experimental results are validated with simulation results. A fair comparison among the MSRF-PLL, FO-LP MSRF-PLL, and FO-LPFO-PI MSRF-PLL is also depicted during grid abnormalities.

KEYWORDS

controller, fractional-order low-pass filter, phase-locked loop, SRF-PLL, synchronization

1 | INTRODUCTION

A vast majority of power electronic-based converter applications involving utility grid interconnections need precise estimation of utility information, particularly phase angle, frequency, and magnitude of the grid voltage.[1] The synchronization of power electronic converters with utility grid has become a crucial factor in managing the growing number of single-phase[2–4] grid-tied power conditioning systems. In general, SRF-PLL is a commonly utilized synchronization mechanism in three-phase grid connection due to its robustness. This technique offers rapid and precise phase/frequency detection and fairly satisfactory dynamic response under ideal grid circumstances. However, the performance

 © 2024 John Wiley & Sons Ltd.

of SRF-PLL deteriorates during adverse grid circumstances.[5] One method of improving the SRF-PLL performance issue is to add a low-pass filter to the control loop and decrease the bandwidth of the SRF PLL at the expense of dynamic performance.6 However, this approach is usually not preferred due to poor dynamic response.

Figure 1 shows the representation of the conventional SRF-PLL employed for single-phase systems, and it is designed comprising three basic parts: (i) a phase detector (PD), (ii) loop filter (LF) that is a PI controller, and (iii) voltage control oscillator (VCO) that monitors the grid voltage and synchronizes its output signal with an input reference signal in both frequency and phase.7,8 The performance of the grid-tied SRF-PLL using the traditional integer order is discussed in various research papers.[9–11] The traditional calculus based on integer order differentiation and integration is popular and convenient due to the absence of suitable and appropriate solution methods for fractional differential equations.[12]

Recently, there has been a lot of study done on fractional differential equations and their applications in different areas like control theory and electrical circuit theory. Many techniques have been used for approximating integrals and fractional derivatives.[13] Today, several applications in physics and engineering employ fractional controllers and employ fractional calculus. Electrical circuits using fractional capacitor (which is proportional to $s^a$where "a" is the arbitrary) are mentioned.[14]

The mathematics that deals with the study and use of integrals and derivatives of any order is defined as fractional calculus.[15]

Several approximations of fractional derivatives in different areas of control theory and PLL are discussed in different papers. It is argued by Stanislawski and Latawiec[16] that normalized finite fractional differences exist.[16] Tseng and some other authors have described numerical realization of differentiators or integrators of fractional variable order.[17,18] Nikhil and his co-authors have presented a cascaded control approach that includes a fractional-order PI (FO-PI) controller for controlling power flow in a renewable-energy-based microgrid connected to utility grid.[19] A servo control technique is suggested for tuning of fractional order PI (FO-PI) controllers with and without time delays.[20] In paper,[15] the authors have proposed a fractional-order analog phase-locked loop (FAPLL) model with fractional-order loop filter and voltage-controlled oscillator and compared the performance of FO-PLL to traditional integer order PLL in terms of capture and locking time response. In paper,[21] fractional-order low-pass filter and fractional-order VCO for FOPLL are used to compute the phase error, output phase, steady-state errors, and stability conditions of FOPLL. A novel fractional-order charge-pump PLL (FOCPLL) with the fractional-order loop filter (FO-LF) is proposed for the overall improvement of the system.[22] A novel PLL algorithm based on fractional-order proportional plus integral (FOPI) controller tuned through genetic algorithm is proposed in Pereira et al.[23] It overcomes the DSOGI PLL's phase peak response by producing a quadrature signal. In paper,[24] $(1 + \alpha)$ Butterworth filter with fractional-order α is presented to study the magnitude-frequency response. In the paper25 various types of "$1 + \alpha$" order Butterworth transfer functions are generated whose cutoff frequencies can be fully tuned. In another paper,26 higher-order fractional filters with fully controlled frequency characteristics are achieved after fitting the filter's magnitude response data with a minimum-phase state-space model. Paper[27] has discussed the design of the fractional-order analog filters and proved that the fractional-order domain adds more degrees of freedom to the filter design process. In paper,[28] fractional-order low pass, band pass, high pass, and all pass filters have been designed using fractional-order impedance in the complex

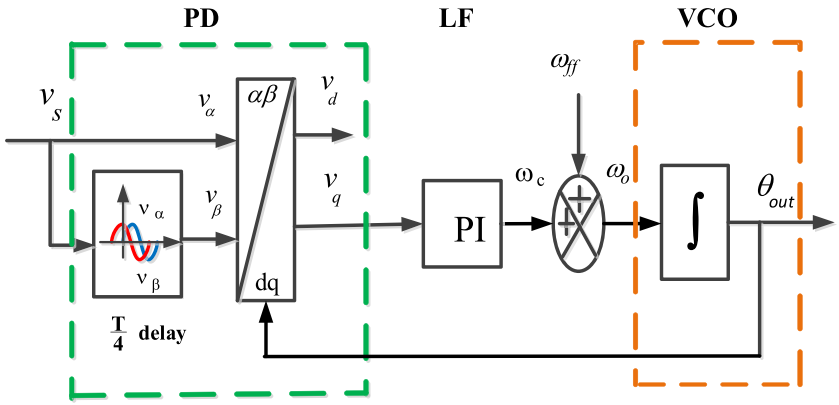


FIGURE 1 Block diagram of conventional SRF-PLL.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

frequency range. It obtains the pole frequencies, half-power frequencies, right-phase frequencies, and quality factor expressions. Paper[29] focuses on developed fractional-order notch filter (FONF) control approach to estimate fundamental active component of current from the distorted load currents and utilizes it to obtain the gating pulses to operate the voltage source converter (VSC). The review of mentioned papers suggests that the use of fractional-order (FO) domain as a PLL for the estimation of frequency and phase angle is interesting and needs elaborate study. Keeping the gap in consideration, this article focuses on modified SRF PLL for single-phase system as well as two new FO SRF-PLL models. The novelty in this article is the introduction of the following:

i. Fractional order in the low-pass filter (FO-LP) of first order
ii. Fractional order in low-pass filter of first order as well as PI controller resulting in (FO-LPFO-PI) MSRF-PLL.

The performance analysis of these two newly developed MSRF-PLLs is studied under grid abnormalities like voltage swell and sag, phase angle change of 20°, frequency variation of + 5 Hz, dc offset, and polluted grid voltage. Thus, three PLLs, viz., (i) MSRF-PLL with low-pass filter having integral order, (ii) MSRF-PLL with low-pass filter having fractional-order (FO-LP), and (iii) MSRF-PLL with two FOs, that is, low-pass filter (FO-LP) and proportional plus integral (FO-PI) are designed and analyzed mathematically in fractional domain. The comparative performance analysis under grid abnormal conditions like 15% voltage swell and sag, phase change of 15°, frequency variation of + 5 Hz, 10% dc offset, and polluted input voltage are rigorously studied.

Theory of fractional calculus emerged with the introduction of regular (integer order) calculus by Leibniz and L'Hopitalin 1695 who proposed half-order derivatives. The idea of fractional calculus further developed in Europe.[30] The base of fractional calculus is the FO differentiations and integrations. It includes works based on Riemann–Liouville definition, the Caputo definition, the Grunwald–Letnikov definition, and the Cauchy integral formula definition and the Cauchy integral formula.[31] Fractional calculus is integrating and differentiating to a noninteger order fundamental operator ${}_{a}\mathrm{D}_{t}^{r'}$ where "a" and "t" are the limits of the operation. The continuous integral-differential operator is defined as

$$ {}_{a}D_{t}^{r'} = \begin{cases} d^{r'}/d^{r'} & R(r') > 0 \\ 1 & R(r') = 0 \\ \int_{0}^{t} d\tau^{-r'} & R(r') < 0 \end{cases} \tag{1} $$

where r' is the operation's order; often, $r' \in R$, but r' may also be a complex number.[19] The two most commonly used definitions in FO calculus are the Grunwald–Letnikov (GL) and the Riemann–Liouville (RL).[13] Letnikov (GL) is defined fractional order as

$$ {}_{a}D_{t}^{r'} f(t) = \lim_{h \to 0} h^{-r'} \sum_{j=0}^{\frac{t-a}{h}} (-1)^{j} \binom{r'}{j} f(t - jh) \tag{2} $$

where $\binom{r'}{j}$ denotes the binomial coefficients; subscripts to the left and right of "D" are the lower bounds and upper bounds in the integral. The value of "r'" can be positive or negative, corresponding to the differentiation and integration, and it is a noninteger. The FO integral according to RL[32] is expressed as

$$ {}_{a}D_{t}^{-r'} f(t) = \frac{1}{\Gamma(r')} \int_{a}^{t} (t - \tau)^{r'-1} f(t) d\tau \tag{3} $$

where $0 < r' < 0$ and "a" is initial value. For "a" $= 0$, the simpler integral notation is ${}_{a}D_{t}^{-r'} f(t)$. The FO differentiation according to RL[13] is expressed as

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$ {}_aD_t^{r/} f(t) = \frac{d^n}{dt^n}\left[{}_aD_t^{-(n-r/)} f(t)\right] = \frac{1}{\Gamma(n-r/)}\frac{d^n}{dt^n}\int_a^t \frac{f(\tau)}{(t-\tau)^{r/-n+1}}d\tau \tag{4} $$

where $(n-1 < r/ \le n)$. The Laplace transformation for the FO differentiation of the RL can be expressed as

$$ \int_0^\infty e^{-st}\,{}_0D_t^{r/} f(t)dt = s^{r/}F(s) - \sum_{k=0}^{n-1} s^k\,{}_0D_t^{r/-k-1} f(t)|_{t=0} \tag{5} $$

Fractional calculus helps to accurately develop models very close to real objects. Recently, its application is observed in many physical and engineering systems. Moreover, FO partial differential equations regulate numerous physical phenomena, including electricity, fluid dynamics, quantum physics, ecological systems, and many other models, within their respective domains of validity.[33] This paper discusses the use of fractional calculus for grid synchronization.

2 | SYSTEM DESCRIPTION

In this section, three mathematical models of modified PLLs are developed. The first model is the MSRF-PLL with integer order. The other two proposed PLLs are (i) FO MSRF-PLL with only one FO capacitor of order a at low-pass filter and (ii) FO MSRF-PLL that have two FO capacitors, both at low-pass filter of order a and FO PI controller of order b.

2.1 | Mathematical modeling of integer order MSRF-PLL

In this subsection, single-phase integer-order MSRF-PLL with low-pass filter of first order $\left(\frac{\omega_{cf}}{s+\omega_{cf}}\right)$ is developed as shown in Figure 2. This low-pass filter of first order is added in phase detection (PD) unit. MSRF-PLL works by introducing 90° phase change (delay of one-fourth of a cycle, T/4) into the input sinusoidal signal (or α signal) to generate the orthogonal signal (β signal) directly for Park transformation (conversion of stationary αβ axis to rotatory dq axis). The voltage referred to quadrature axis (v_q) is set to zero value so that output of MSRF-PLL locks into the grid utility voltage's phase angle (v_s*). The noise associated with v_q voltage traverses through the low-pass filter of first order that is further passed through filter of the loop (LF) and voltage-controlled oscillator (VCO) unit. The output of LF is the controlled frequency output (ω_c) that along with the nominal frequency ($\omega_{ff} = 2\pi f$) gives predicted frequency (ω_o) of grid voltage. The estimated phase angle θ_{out} of the input signal is derived through VCO by the integration of the calculated frequency (ω_o). All the control loops or the loop dynamics of MSRF-PLL are defined using integer-order transfer functions. The sinusoidal input grid voltage (v_s) is equivalent to v_α, and its $\frac{\pi}{2}$ time delay signal (v_β) is transformed into the dq voltage by the following equations where $T = 2\pi$.

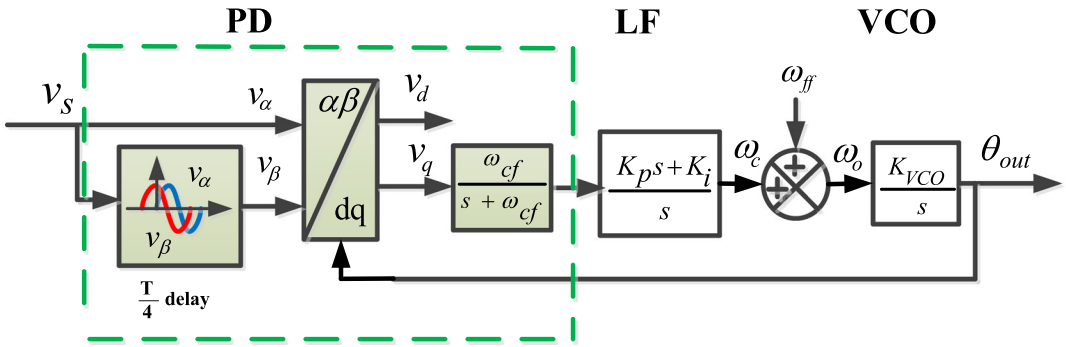


FIGURE 2 Block diagram of single phase MSRF-PLL in integral order domain.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\left.\begin{aligned} v_s &= V_s sin\omega_o t = V_s sin\theta_{out} \\ sin\left(\omega_o t + \frac{t}{4}\right) &= cos(\omega_o t) = cos\theta_{out} \\ v_{s\alpha} &= V_s sin\theta_{out} \\ v_{s\beta} &= V_s sin\left(\omega_o t + \frac{T}{4}\right) = V_s cos\theta_{out} \end{aligned}\right\} \tag{6}$$

Park transformation is applied to v_α and v_β from (6) to get v_d and v_q as given in (7) and (8).

$$\begin{bmatrix} v_{sd} \\ v_{sq} \end{bmatrix} = \begin{bmatrix} sin\theta & -cos\theta \\ cos\theta & sin\theta \end{bmatrix} \begin{bmatrix} v_{s\alpha} \\ v_{s\beta} \end{bmatrix} \tag{7}$$

$$\left.\begin{aligned} v_{sd} &= V_{s\alpha}\sin\theta - V_{s\beta}\cos\theta \\ v_{sq} &= V_{s\alpha}\cos\theta + V_{s\beta}\sin\theta \end{aligned}\right\} \tag{8}$$

where θ_{out} is the estimated angle in degree (°) given by transfer function $\theta_{out} = \frac{K_{VCO}V_c(s)}{s}$ which can be interpreted as the integral of control voltage; K_{VCO} is the VCO gain; V_c is the controlled voltage obtained as output from PI controller; $\omega_o = \omega_c + \omega_{ff}$ is the calculated frequency in Hz; v_s is input signal voltage in volts; and ω_{ff} is the nominally stated frequency; $\omega_{ff} = 100\pi$rad/s. The amplitude of v_s is calculated as $A_t = \sqrt{(v_{sd})^2 + (v_{sq})^2}$. The phase angle error, θ_{error}, of fundamental voltage, v_s, is computed from v_q at PCC of grid utility converters, and this v_q is passed through a low-pass filter (LPF) of first order to exclude high-amplitude steady-state oscillations in the calculated phase and frequency. The transfer function of LPF is expressed as

$$Tf_{LPF}(s) = \frac{\omega_{cf}}{s + \omega_{cf}} \tag{9}$$

where ω_{cf} is the cutoff frequency and the order of "s" is an integer. The PD output is a phase error passing through the loop filter (LF). The LF commonly designed for synchronization applications is a proportional and integral (PI) controller. The transfer function of PI controller is given as

$$Tf_{LF}(s) = \frac{K_p s + K_i}{s} \tag{10}$$

where K_p and K_i are the proportional and integral gains, respectively. A controlled voltage,V_c, that is proportional to controlled angular frequency, ω_c, is an output from PI controller that in summation with nominal frequency ω_{ff} is passed through VCO whose transfer function is given by

$$Tf_{VCO}(s) = \frac{K_{VCO}}{s} \tag{11}$$

which after integrating the estimated phase angle θ_{out} is obtained as the output from VCO. The transfer function of the open-loop system, also known as loop gain (LG) of the system, is shown in Figure 2 and expressed as

$$\begin{aligned} Tf_{LG}(s) &= [Tf_{PD}(s)]\,[Tf_{LF}(s)]\,[Tf_{VCO}(s)] \\ &= \left(\frac{K_{PD}\omega_{cf}}{s + \omega_{cf}}\right)\left(\frac{K_p s + K_i}{s}\right)\left(\frac{K_{VCO}}{s}\right) \\ &= \left(\frac{Ks + K^{/}}{s^3 + s^2\omega_{cf}}\right) \end{aligned} \tag{12}$$

where $K = K_{PD}K_{VCO}K_p\omega_{cf}$ and $K^{/} = K_{PD}K_{VCO}K_i\omega_{cf}$. The closed-loop transfer function for integer order MSRF-PLL is given by

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$H(s) = \frac{\theta_{output}}{\theta_{input}} = \frac{Tf_{LG}(s)}{1 + Tf_{LG}(s)} = \left(\frac{Ks + K^{/}}{s^3 + \omega_{cf} s^2 + Ks + K^{/}} \right) \tag{13}$$

The error response transfer function for MSRF-PLL is expressed as

$$E(s) = \frac{\theta_{error}}{\theta_{input}} = \frac{1}{1 + Tf_{LG}(s)} = \left(\frac{s^3 + \omega_{cf} s^2}{s^3 + \omega_{cf} s^2 + Ks + K^{/}} \right) \tag{14}$$

where $\theta_{input}, \theta_{error}, \theta_{output}$, are, respectively, the Laplace transformation for phase angle of input, error, and output signals.

2.2 | Mathematical modeling of FO MSRF-PLL

Two new configurations of FO MSRF-PLL are displayed in Figures 3 and 4, and their analysis is discussed in detail. The first configuration of FO MSRF-PLL algorithm is designed by utilizing a FO-LP filter of coefficient a as shown in Figure 3.

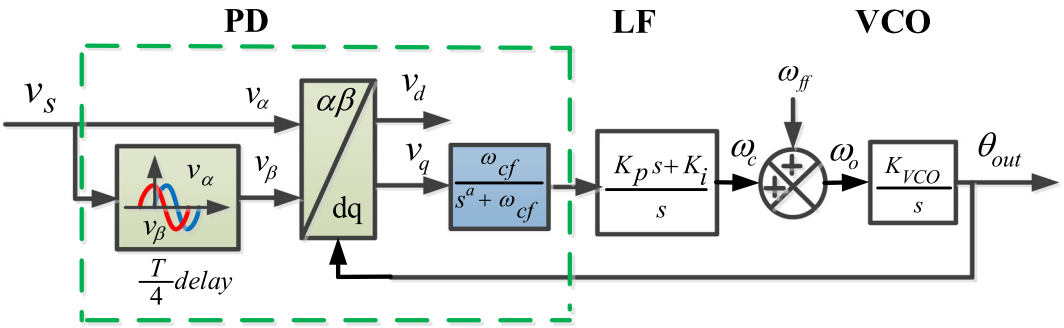


FIGURE 3 Diagram of MSRF-PLL with FO low-pass filter (FO-LP).

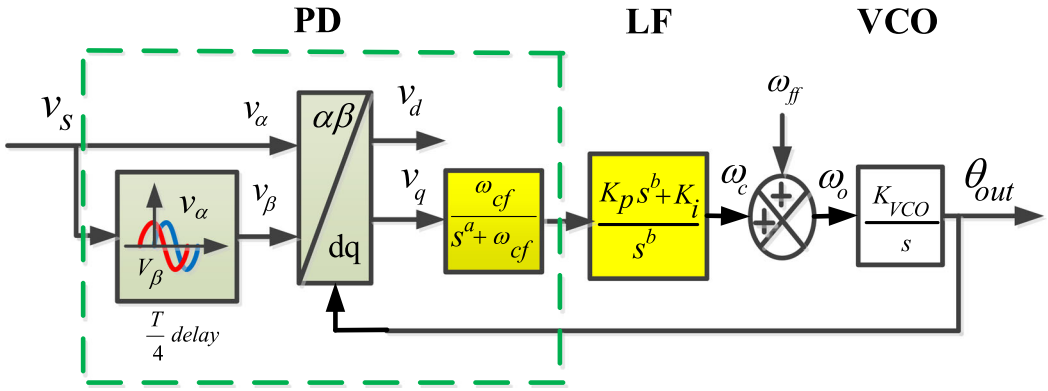


FIGURE 4 Block diagram of MSRF-PLL with fractional-order low-pass filter (FO-LP) and fractional-order proportional plus integrator (FO-PI).

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The second algorithm is designed considering FO MSRF-PLL which has both FO-LP filter of order a and FO-PI of fractional order b as shown in Figure 4. The FO MSRF-PLL is made fractional using FO capacitance and is given by $Z(s) = Cs^{\alpha} \Rightarrow Z(j\omega)$ in Laplace domain with $\alpha = -1$. The transfer functions of the two configurations designed in Figures 3 and 4 are discussed here in detail. The transfer functions of FO-LP filter having FO a $(0 < a < 2)$ and FO-PI having FO b $(0 < b < 2)$ can be expressed as

$$Tf_{FO-LP}(s) = \frac{\omega_{cf}}{s^a + \omega_{cf}} \tag{15}$$

$$Tf_{FO-PI}(s) = K_p + \frac{K_i}{s^b} \tag{16}$$

The transfer function of the VCO unit remains unchanged as given by Equation (11).

2.2.1 | Mathematical model of fractional-order MSRF-PLL with only fractional-order low-pass filter (FO-LP MSRF-PLL)

In this subsection, the mathematical model of various transfer functions for open loop, closed loop, and error response of the FO MSRF-PLL having only low-pass filter (FO-LP MSRF-PLL) is developed. The transfer function for open-loop gain FO-LP MSRF-PLL is expressed as

$$\begin{aligned} Tf_{FOLG}(s) &= K_{PD}[Tf_{FO-LP}(s)][Tf_{LF}(s)][Tf_{VCO}(s)] \\ &= \left(\frac{K_{PD}\omega_{cf}}{s^a + \omega_{cf}}\right)\left(\frac{K_p s + K_i}{s}\right)\left(\frac{K_{VCO}}{s}\right) \\ &= \left(\frac{K_{PD}K_P K_{VCO}\omega_{cf} s + K_{PD}K_i K_{VCO}\omega_{cf}}{s^{2+a} + \omega_{cf} s^2}\right) \\ &= \left(\frac{Ks + K^{/}}{s^{2+a} + \omega_{cf} s^2}\right) \end{aligned} \tag{17}$$

where $K = K_{PD}K_{VCO}K_p\omega_{cf}$ and $K^{/} = K_{PD}K_{VCO}K_i\omega_{cf}$. Closed-loop transfer function of FO-LP MSRF-PLL is expressed as

$$\begin{aligned} H_{FO-LP}(s) &= \left(\frac{\theta_{output}}{\theta_{input}}\right) = \left(\frac{Tf_{FOLG}(s)}{1 + Tf_{FOLG}(s)}\right) \\ &= \left(\frac{K_{PD}K_{VCO}K_p\omega_{cf} s + K_{PD}K_{VCO}K_i\omega_{cf}}{s^{2+a} + \omega_{cf} s^2 + K_{PD}K_{VCO}K_p\omega_{cf} s + K_{PD}K_{VCO}K_i\omega_{cf}}\right) \\ &= \left(\frac{Ks + K^{/}}{s^{2+a} + \omega_{cf} s^2 + Ks + K^{/}}\right) \end{aligned} \tag{18}$$

where $K = K_{PD}K_{VCO}K_p\omega_{cf}$ and $K^{/} = K_{PD}K_{VCO}K_i\omega_{cf}$.

The error response transfer function of the FO-LP MSRF-PLL is expressed as

$$\begin{aligned} E_{FO-LP}(s) &= \left(\frac{\theta_{error}}{\theta_{input}}\right) = \left(\frac{1}{1 + Tf_{FOLG}(s)}\right) \\ &= \left(\frac{s^{2+a} + \omega_{cf} s^2}{s^{2+a} + \omega_{cf} s^2 + Ks + K^{/}}\right) \end{aligned} \tag{19}$$

where $K = K_{PD}K_{VCO}K_p\omega_{cf}$, $K^{/} = K_{PD}K_{VCO}K_i\omega_{cf}$ and $\theta_{input}, \theta_{error}, \theta_{output}$, are, respectively, the Laplace transformation for phase angle of input signal, error signal, and output signal.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2.2 | Mathematical model of fractional-order MSRF-PLL with both fractional-order low-pass filter (FO-LP) and loop filter (FO-PI)

In this subsection, FO-LPFO-PI MSRF-PLL has been modeled considering both fractional low-pass filter and fractional loop filter, which is a PI controller. The open-loop gain transfer function of FO-LPFO-PI MSRF-PLL is expressed as

$$\begin{aligned} Tf_{LGFOLPFOPI}(s) &= K_{PD}(Tf_{FO-LP})(Tf_{FO-PI})(Tf_{VCO}) \\ &= \left(\frac{K_{PD}\omega_{cf}}{s^a+\omega_{cf}}\right)\left(\frac{K_p s^b + K_i}{s^b}\right)\left(\frac{K_{VCO}}{s}\right) \\ &= \left(\frac{K_{PD}K_PK_{VCO}\omega_{cf}s^b + K_{PD}K_{VCO}K_i\omega_{cf}}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}}\right) \\ &= \left(\frac{Ks^b + K^/}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}}\right) \end{aligned} \tag{20}$$

where $K = K_{PD}K_{VCO}K_P\omega_{cf}$ and $K^/ = K_{PD}K_{VCO}K_i\omega_{cf}$.

The closed-loop transfer function of FO-LPFO-PI MSRF-PLL is expressed as

$$\begin{aligned} H_{FOLPFOPI}(s) &= \left(\frac{\theta_{output}}{\theta_{input}}\right) = \left(\frac{Tf_{LGFOLPFOPI}(s)}{1+Tf_{LGFOLPFOPI}(s)}\right) \\ &= \left(\frac{K_{PD}K_pK_{VCO}\omega_{cf}s^b + K_{PD}K_iK_{VCO}\omega_{cf}}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}+K_{PD}K_PK_{VCO}\omega_{cf}s^b + K_{PD}K_iK_{VCO}\omega_{cf}}\right) \\ &= \left(\frac{Ks^b + K^/}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}+Ks^b + K^/}\right) \end{aligned} \tag{21}$$

where $K = K_{PD}K_PK_{VCO}\omega_{cf}$and $K^/ = K_{PD}K_iK_{VCO}\omega_{cf}$.

The transfer function for error response of FO-LPFO-PI MSRF-PLL is expressed as

$$\begin{aligned} E_{FOLPFOPI}(s) &= \left(\frac{\theta_{error}}{\theta_{input}}\right) = \left(\frac{1}{1+Tf_{LGFOLPFOPI}(s)}\right) \\ &= \left(\frac{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}+K_{PD}K_PK_{VCO}\omega_{cf}s^b + K_{PD}K_iK_{VCO}\omega_{cf}}\right) \\ &= \left(\frac{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}+Ks^b + K^/}\right) \end{aligned} \tag{22}$$

where $K = K_{PD}K_PK_{VCO}\omega_{cf}$, $K^/ = K_{PD}K_iK_{VCO}\omega_{cf}$, and $\theta_{input}, \theta_{error}, \theta_{output}$, are, respectively, the phase-angle Laplace transformation for the signals of input, error, and output.

3 | ANALYSIS OF FRACTIONAL-ORDER MSRF-PLLS

This section focuses on the stability analysis of fractional MSRF-PLL using Bode plots. The open-loop and closed-loop transfer functions of FO-LP MSRF-PLL and FO-LPFO-PI MSRF-PLL are already described in Equations (17), (18), (20), and (21). It is well known that if the phase margin and gain margin are both positive, the FO system is considered stable.[13] The reciprocal of the frequency response amplitude in which the phase is -180° at crossover frequency (ω_{pc}) is the gain margin (Gm). It is measured in dB and expressed as

$$\mathrm{Gm} = -20\log\,|\,Tf_{LG}(j\omega_\pi)\,|\,\mathrm{dB}. \tag{23}$$

where $Tf_{LG}(j\omega_\pi)$ is the transfer function for the open-loop system. The frequency where either of the two $\pm 180^\circ$ axis crosses the open-loop locus is called as the phase crossover frequency (ω_{pc}).

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The phase margin is defined as the difference in phase between 180° and gain crossover and is measured in degrees. It is expressed as $\text{Pm} = \arg\,[Tf_{LG}(j\omega_{\pi})] + \pi$ (24)

The crossover frequency of the open-loop plot along 0-dB axis is termed as the gain crossover frequency (ω_{gc}).

The Bode diagrams (Figure 5 and Table 1) are analyzed to optimize the value of the fractional parameter a for stable FO-LP MSRF-PLL system. Figure 5A,B shows the Bode plots of open-loop and closed-loop transfer functions of FO-LP MSRF-PLL for various values of a ($0.2 \leq a \leq 1.7$) with step increment of 0.3 in a.

Table 1 gives the values of gain margins with its corresponding phase crossover frequencies and phase margins with its corresponding gain crossover frequencies for open-loop and closed-loop transfer functions of FO-LP MSRF-PLL for different values of a. From Table 1, it is evident that the phase margin is positive and almost constant for $a = 0.2$ to 1.7 for both open-loop and closed-loop system. However, for $a = 0.2$ to 0.8, the system is unstable as Gm is negative for open loop. The closed loop shows Gm is infinite with undefined phase crossover frequency. On varying a from 1.1 to 1.7, it is observed that the Gm is positive, but it decreases considering both open-loop and closed-loop systems. Hence, it can be concluded from Table 1 and Figure 5A,B that the designed FO-LP MSRF-PLL gives best performance with optimized value of a as 1.1.

The Bode diagrams (Figures 6 and 7 and Tables 2 and 3) are analyzed to optimize the values of the fractional order parameters a and b for FO-LPFO-PI MSRF-PLL system. Figure 6A,B gives the open-loop and closed-loop transfer function of FO-LPFO-PI MSRF-PLL for fixed $b = 1$ and variable a ($0.2 \leq a \leq 1.7$) having step change of 0.3. Table 2 gives information of the values of gain margin with corresponding phase crossover frequencies and phase margins with corresponding gain crossover frequencies. For $0.2 \leq a \leq 0.8$, the Gm is negative for open-loop and infinite for closed-loop system. From Table 2 and Bode diagram shown in Figure 6A,B, it is observed that for $a = 1.1$ to 1.7, the Gm

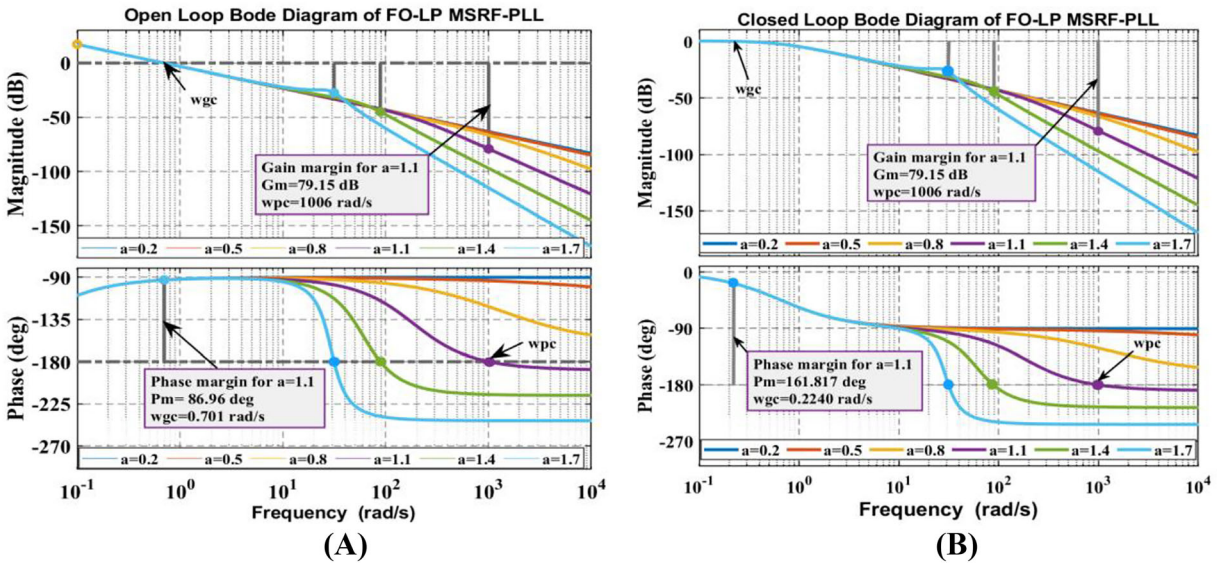


FIGURE 5 Bode diagram of (A) open-loop and (B) closed-loop transfer functions of FO-LP MSRF-PLL.

TABLE 1 Values of gain margin, phase margin, and crossover frequencies for different values of a for FO-LP MSRF-PLL.

	Value of gm (dB)		Value of ω_{pc} (rad/s)		Value of pm (°)		Value of ω_{gc} (rad/s)	
Value of a	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**
$a = 0.2$	− 179.9	Infinite	5.0192e−06	Undefined	87.0225	161.789	0.6989	0.2238
$a = 0.5$	− 184.9	Infinite	3.7692e−06	Undefined	86.9699	161.780	0.6996	0.2242
$a = 0.8$	− 185.1	Infinite	3.7221e−06	Undefined	86.9505	161.796	0.7004	0.2242
$a = 1.1$	**79.16**	**79.15**	**1005.8**	**1006**	**86.9620**	**161.817**	**0.7011**	**0.2240**
$a = 1.4$	44.84	44.79	88.7913	88.7913	86.9963	161.836	0.7017	0.2238
$a = 1.7$	27.21	26.82	31.4975	31.4978	87.0420	161.847	0.7020	0.2237

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

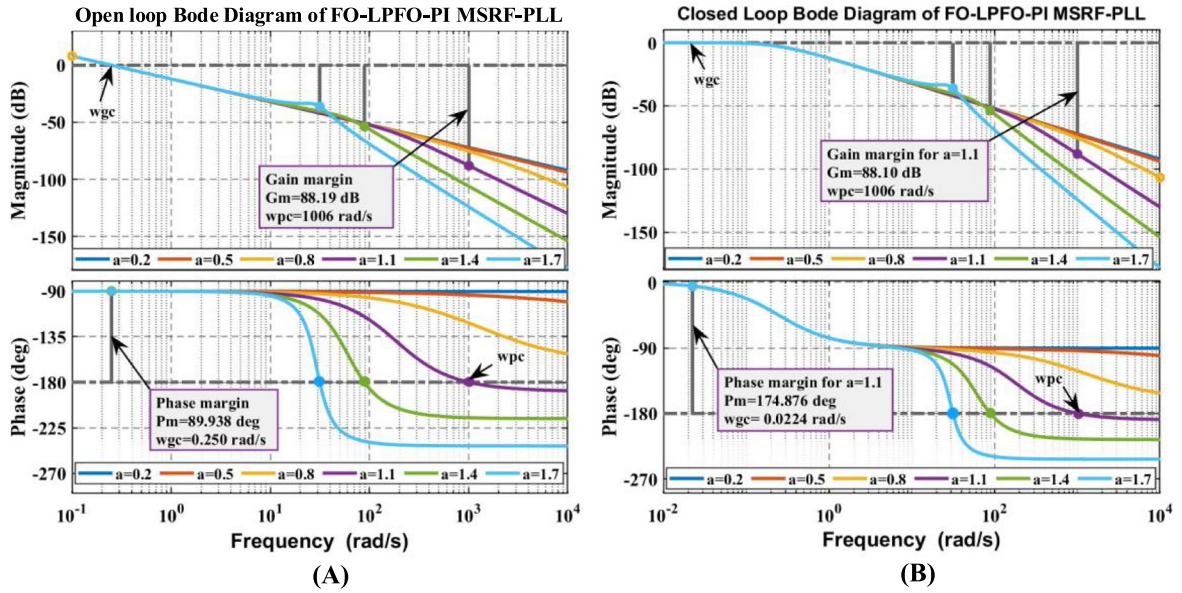


FIGURE 6 Bode diagram of (A) open-loop and (B) closed-loop transfer function of FO-LPFO-PI MSRF-PLL for fixed $b = 1$ and different values of a.

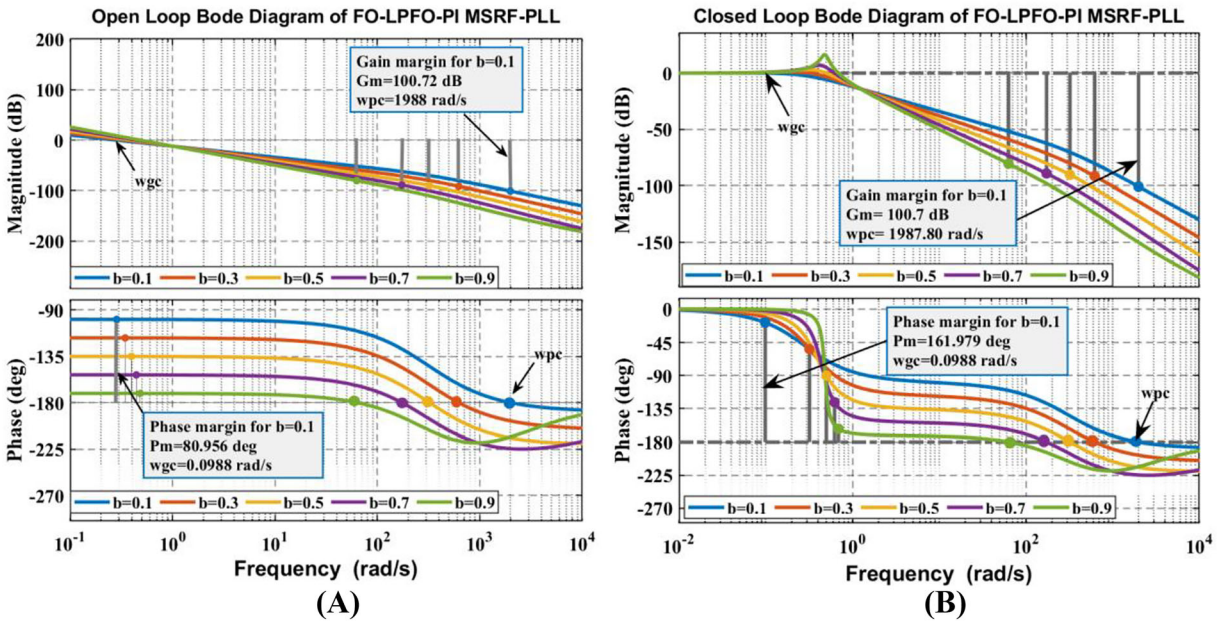


FIGURE 7 (A,B) Bode diagram of open-loop and closed-loop transfer function FO-LP-PI MSRF-PLL for fixed $a = 1$ with different values of b.

decreases with increase in a, but there is minimal increment in Pm. Hence, it can be inferred from Table 2 that for $a = 1.1$, the system remains stable showing highest positive value of Gm and Pm, and this is the optimized choice for parameter a.

Figure 7A,B shows Bode diagrams of the open-loop and closed-loop transfer functions of FO-LPFO-PI MSRF-PLL for fixed $a = 1$ and variable b ($0.1 \leq b \leq 0.9$) having step change of 0.2. Table 3 shows different values of Gm with corresponding phase crossover frequencies and with corresponding gain crossover frequencies. For both the open-loop and closed-loop system, it is observed that with an increase in b ($0.1 \leq b \leq 0.9$), both Gm and Pm decrease. Hence, it can be analyzed from Figure 7A,B and Table 3 that FO-LPFO-PI

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Values of gain margin, phase margin, and crossover frequencies for different values of a with fixed b as 1 for FO-LPFO-PI MSRF-PLL.

	Value of gm (dB)		Value of phase crossover frequency ω_{pc} (rad/s)		Value of pm (°)		Value of gain crossover frequency ω_{gc} (rad/s)	
Value of a	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**
$a = 0.2$	− 125.00	Infinite	3.6989e−06	Undefined	89.934	174.849	0.249	0.0225
$a = 0.5$	− 125.20	Infinite	3.6962e−06	Undefined	89.913	174.859	0.250	0.0224
$a = 0.8$	− 125.21	Infinite	3.6960e−06	Undefined	89.920	174.871	0.250	0.0224
$a = 1.1$	**88.19**	**88.10**	**1006**	**1006**	**89.938**	**174.876**	**0.250**	**0.0224**
$a = 1.4$	83.79	53.77	88.8267	88.8258	89.956	174.878	0.250	0.0224
$a = 1.7$	36.16	36.02	31.5083	31.5082	89.969	174.879	0.250	0.0224

TABLE 3 Values of gain margin, phase margin, and crossover frequencies for different values of b and $a = 1$ for FO-LPFO-PI MSRF-PLL.

	Value of gm (dB)		Value of ω_{pc} (rad/s)		Value of pm (°)		Value of ω_{gc} (rad/s)	
Value b	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**	**Open loop**	**Closed loop**
$b = 0.1$	**100.718**	**100.72**	**1987.8**	**1987.80**	**80.956**	**161.979**	**0.2838**	**0.0988**
$b = 0.3$	91.528	91.53	621.362	621.362	62.956	125.923	0.3444	0.3201
$b = 0.5$	90.292	90.29	322.134	322.134	44.953	89.892	0.3970	0.5005
$b = 0.7$	89.143	89.14	173.073	173.073	26.948	53.854	0.4425	0.6219
$b = 0.9$	80.522	80.52	62.9963	62.996	8.941	17.8374	0.4821	0.6900

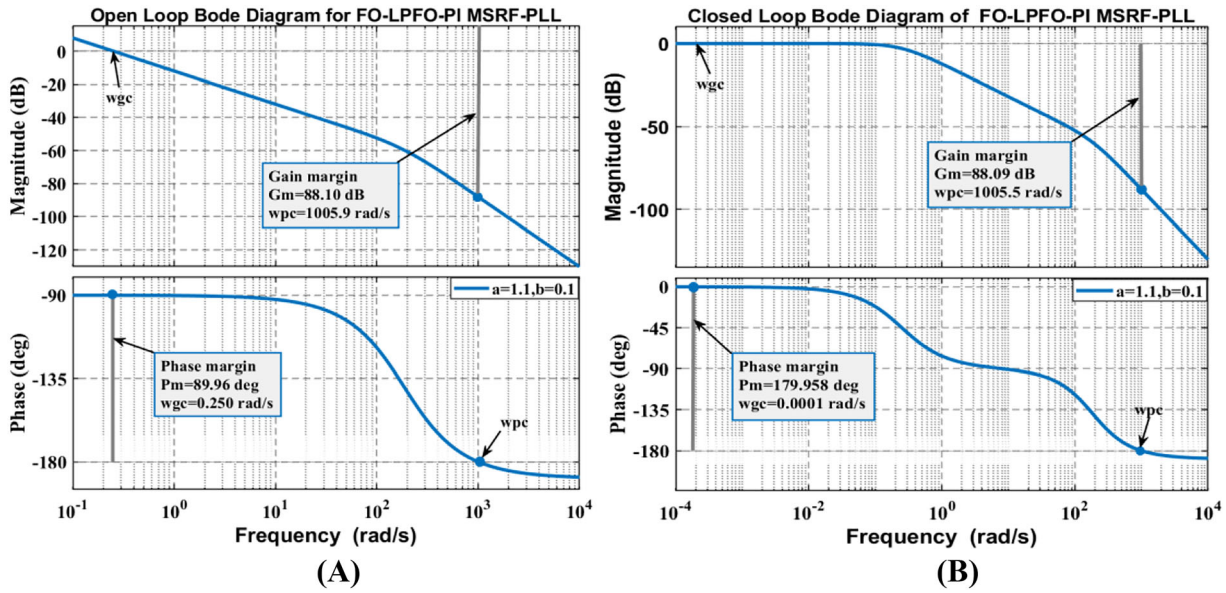


FIGURE 8 (A) Open-loop and (B) closed-loop transfer function of FO-LP-PIMSRF-PLL for $a = 1.1$ and $b = 0.1$.

MSRF-PLL system is most stable for $b = 0.1$ when a is kept fixed at 1. Thus, the optimized values for FO-LPFO-PI MSRF-PLL system are selected as $a = 1.1$ and $b = 0.1$ respectively.

Figure 8A,B shows the final Bode diagrams for the open-loop and closed-loop transfer functions for FO-LPFO-PI MSRF-PLL corresponding to selected optimized values of $a = 1.1$ and $b = 0.1$. The open loop gives a gain margin, Gm = 88.10 dB with phase crossover frequency (ω_{pc}) at 1005.9 rad/s and corresponding phase margin is Pm = 89.96°

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 4 Summarized results for FO-LP MSRF-PLL and FO-LPFO-PI MSRF-PLL.

Types of MSRF-PLL		Open-loop transfer function	Closed-loop transfer function
FO-LP MSRF-PLL		$Tf_{FOLG} = K_{PD}Tf_{FO-LP}(s)Tf_{LF}(s)Tf_{VCO}(s) = \left(\frac{Ks+K^{/}}{s^{2+a}+\omega_{cf}s^{2}}\right)$	$H_{FO-LP}(s) = \left(\frac{Tf_{FOLG}(s)}{1+Tf_{FOLG}(s)}\right) = \left(\frac{Ks+K^{/}}{s^{2+a}+\omega_{cf}s^{2}+Ks+K^{/}}\right)$
	At optimized value of $a = 1.1$	**79.16 dB**	**79.15 dB**
	Gm (dB)	**86.9620°**	**161.817°**
	Pm (°)		
FO-LPFO-PI MSRF-PLL		$Tf_{LGFOLPFOPI}(s) = K_{PD}(Tf_{FO-LP})(Tf_{FO-PI})(Tf_{VCO}) = \left(\frac{Ks^{b}+K^{/}}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}}\right)$	$H_{FOLPFOPI}(s) = \left(\frac{Tf_{LGFOLPFOPI}(s)}{1+Tf_{LGFOLPFOPI}(s)}\right) = \left(\frac{Ks^{b}+K^{/}}{s^{(1+a+b)}+\omega_{cf}s^{(1+b)}+Ks^{b}+K^{/}}\right)$
	At optimized value of $a = 1.1$ and $b = 0.1$	**88.10 dB**	**88.09 dB**
	Gm (dB)	**89.96°**	**179.958°**
	Pm (°)		

with gain crossover frequency (ω_{gc}) at 0.25 rad/s. The closed-loop analysis shows Gm of 88.09 dB with phase crossover frequency (ω_{pc}) at 1005.5 rad/s and Pm = 179.958°. The FO-LPFO-PI MSRF-PLL is found stable for the selected and optimized values of a and b as both phase and gain margin are positive.

Table 4 discusses the open-loop and closed-loop transfer functions of FO-LP MSRF-PLL with fractional parameter a as derived in Equations (17) and (18), and relevant Bode diagrams are presented in Figure 5. FO-LPFO-PI MSRF-PLL is designed with two fractional parameters a and b as per Equations (20) and (21) for its open- and closed-loop response. The Bode diagrams for FO-LPFO-PI MSRF-PLL are plotted in Figures 6 and 7 that show how parameters a and b can be optimized. The procedure followed for obtaining optimization of a and b involves fixing one fractional parameter while varying the other parameter. Figure 6 shows the Bode plots for FO-LPFO-PI MSRF-PLL with variable a and fixed $b = 1$. Figure 7 shows the Bode plot for variable b with fixed $a = 1$. The final open-loop and closed-loop transfer functions and Gm and Pm values are also indicated for the proposed FO-LPFO-PI MSRF-PLL (Table 5).

4 | SIMULATION AND EXPERIMENTAL RESULTS

The proposed FO-LP MSRF-PLL and FO-LPFO-PI MSRF-PLL are now tested using single-phase programmable grid voltage supply. The simulation is run with sampling time, Ts = 50e−6 s under grid abnormalities during disturbances introduced from time interval from 0.2 to 0.4 s. The voltage sag and swell condition causing amplitude variation may be due to voltage drop, change in load, relay malfunctioning, or transformer tap change in power system. Polluted supply is due to presence of nonlinear loads at the consumer end such as power electronic devices.

Various abnormal conditions of grid-tied supply voltage like voltage sag, voltage swell, polluted voltage, phase angle jump, frequency change, and dc offset are simulated, and the performance of the fractional PLLs for the proposed system is investigated. The fractional-order modeling and control (FOMCON) toolbox for MATLAB/Simulink is used for the performance analysis of the simulation results of the proposed system models. The experimental results of the proposed models are obtained from the laboratory setup consisting of a digital signal oscilloscope (DSO), Infinite Vision DSO-X-2014A, PC and real-time simulator.

4.1 | Dynamic performance under steady-state condition

The value of the parameters considered in the PLLs for simulation is given in Appendix A. Both the simulation and experimental performance analysis of the proposed system models are carried out with the real fractional parameter values for $a = 1.1$ and $b = 0.1$. The dynamic performance analysis of the proposed FO-LPMSRF-PLL and

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 5 Comparative performance analysis of MSRF-PLL, FO-LP MSRF-PLL, and FO-PLFO-PI MSRF-PLL.

Testing conditions	MSRF-PLL	FO-LP MSRF-PLL	FO-LPFO-PI MSRF-PLL
Under 30% voltage sag			
Frequency overshoot	Larger peak overshoot of 51.782 Hz	Smaller peak overshoot of 50.99 Hz	Least peak overshoot of 50.28 Hz
Settling time	slow response (0.03 s)	Slower response (0.043 s)	fastest response (0.023 s)
Under 40% voltage swell			
Frequency overshoot	Larger peak overshoot of 53.42 Hz	Smaller with peak overshoot of 51.8 Hz	Smallest peak overshoot of 50.6 Hz
Settling time	Slow (0.04 s)	Sluggish (0.05 s)	Faster (0.02 s)
Under polluted grid supply			
Frequency overshoot	Largest peak overshoot of 2.95 Hz	Smaller peak overshoot of 1.305 Hz	Smallest peak overshoot of 0.44 Hz
Settling time	Slower (0.032 s)	Sluggish (0.033 s)	Faster (0.02 s)
Phase angle shift of 15°			
Frequency overshoot	Largest peak overshoot of 55.42 Hz	Smaller peak overshoot of 54 Hz	Smallest peak overshoot of 51.95 Hz
Settling time	Slow (0.051 s)	Sluggish (0.06 s)	Faster (0.05 s,)
+5 Hz Frequency change			
Frequency overshoot	Largest peak oscillation of 58 Hz	Smaller peak overshoot of 55.9 Hz	Smallest peak overshoot of 55.3 Hz
Settling time	Slower (0.049 s)	Faster (0.05 s)	Fastest (0.048 s)
Presence of dc offset			
Response in frequency	Large oscillations	Small oscillations	Minimum oscillations

FO-LPFO-PIMSRF-PLL is carried out by analyzing the grid parameters like input voltage supply (v_s) in volts, estimated frequency of the system in hertz), phase angle in degrees, and the amplitude (A_t) of source voltage at the point of common coupling of the system in volts, which is expressed as $A_t = \sqrt{v_{s\alpha}^2 + v_{s\beta}^2}$. Here, $v_{s\alpha}$ and $v_{s\beta}$ are the alpha and beta components of supply voltage.

The simulated and the experimental results of FO-LP MSRF-PLL for the dynamic performance analysis under various grid abnormalities are shown in Figure 9 (i) and (ii). Figure 9 (i) (a) gives the simulated dynamic performance at 30% input voltage sag ($v_s = 227.5$ V) during time interval of 0.2 to 0.4 s, and after 0.4 s, the normal condition is restored. The analysis is observed by studying "v_s" (V), estimated frequency (Hz), phase angle (°), and voltage amplitude A_t (V), and Figure 9 (ii) (a) validates its performance analysis experimentally. The estimated frequency is maintained at fundamental frequency while input voltage amplitude is decreased by 97.5 V (30% of v_s) during disturbance time interval. There is no change in the calculated phase angle. The value of the magnitude A_t diminishes during voltage sag conditions and is correctly estimated both in simulation and hardware results. Figure 9 (i) (b) gives the simulated dynamic performance under 40% input voltage swell ($v_s = 455$ V) during time interval of 0.2 to 0.4 s. Figure 9 (ii) (b) verifies the same case experimentally. From the analysis, small variations in estimating fundamental frequency can be observed with increased input voltage amplitude during swell condition. There is no change in the estimated phase angle, and this is verified from simulation and experimental results.

Figure 9 (i) (c) gives the simulated dynamic performance under polluted grid conditions when harmonic content is added to input signal during time interval of 0.2 to 0.4 s. The supply voltage v_s which is sinusoidal becomes distorted (nonsinusoidal) during the disturbed interval due to harmonic content. Magnitude A_t also shows some distortions in both simulation and experimental results. However, the frequency and phase angle are well tracked in Figure 9 (i) (c) which is verified experimentally by Figure 9 (ii) (c).

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

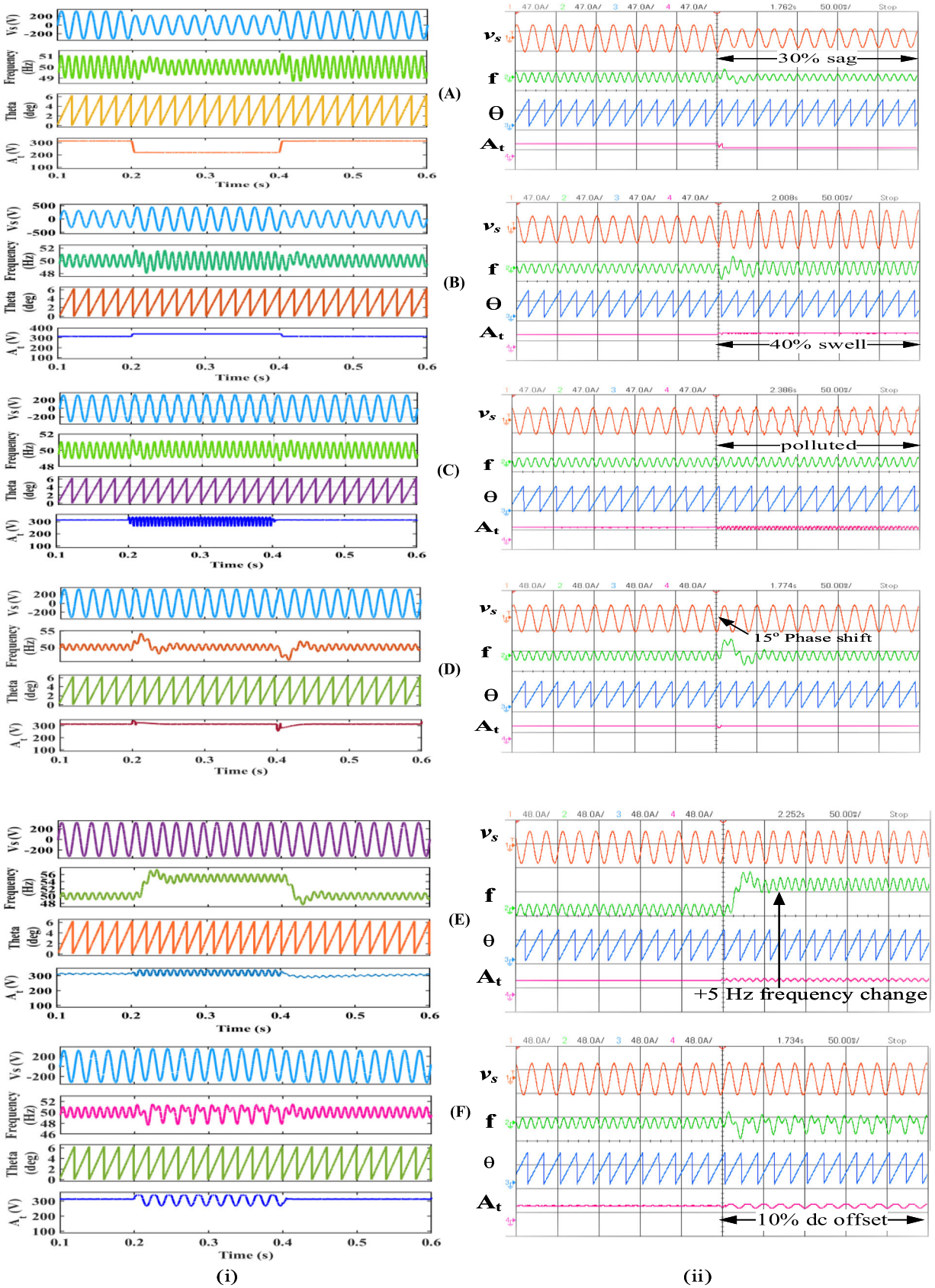


FIGURE 9 (i) Simulated and (ii) experimental results showing dynamic performances of FO-LPMSRF-PLL under unbalance grid conditions for (a) 15% voltage sag, (b) 15% voltage swell, (c) polluted voltage, (d) 15° phase jumps, (e) + 5 Hz frequency jump, and (f) 10% dc offset.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 9 (i) (d) gives the simulated dynamic performance under 15° phase shift. The estimated frequency shows a peak overshoot of 54.32 Hz at 0.213 s. The overshoot in frequency settles down at 0.26 s. Figure 9 (ii) (d) verifies the simulated observation experimentally.

Figure 9 (i) (e) gives the simulated dynamic performance under +5 Hz change in frequency during time interval of 0.2 to 0.4 s. The frequency change leads to a peak oscillation of 55.92 Hz at 0.23 s and settles down at 0.25 s, and Figure 9 (ii) (e) upholds the simulation results experimentally. Thus, the developed FO-LP MSRF-PLL is able to track the frequency of 55 Hz within two cycles, and this result is validated by simulation as well asexperimental studies.

Figure 9 (i) (f) presents the simulated dynamic performance under 10% dc offset during time interval of 0.2 to 0.4 s. The presence of this offset is predicted in v_s and A_t. The distortion in frequency observed through phase angle is tracked faithfully. Figure 9 (ii) (e) proves the results experimentally under dc offset case.

The simulated and experimental dynamic performance of FO-LPFO-PI MSRF-PLL under various grid abnormal conditions like voltage sag, voltage swell, polluted voltage, phase jumps, step change of frequency, and dc offset are shown in Figure 10 (i) and (ii), respectively.

Figure 10 (i) (a) gives the simulated dynamic performance under 30% input voltage sag ($v_s = 227.5$ V) during time interval of 0.2 to 0.4 s. The analysis is performed by observing plots of v_s (V), estimated frequency (Hz), phase angle (°), and voltage amplitude A_t (V), while Figure 10 (ii) (a) confirms the performance analysis experimentally. The estimated frequency and phase angle remain unaffected though input voltage amplitude is decreased during disturbed time interval.

Figure 10 (i) (b) gives the simulated dynamic performance under 40% input voltage swell ($v_s = 455$ V) during interval of 0.2 to 0.4 s. The frequency is well tracked, while the input voltage amplitude is increased by 40% of the input supply voltage during disturbance period. There is no change in the estimated phase angle, and Figure 10 (ii) (b) validates this result experimentally.

Figure 10 (i) (c) gives the simulated dynamic performance under polluted input voltage supply during time interval of 0.2 to 0.4 s. The supply voltage is nonsinusoidal during 0.2–0.4 s owing to harmonic content. The computed input voltage magnitude A_t shows some oscillations, while frequency and phase angle are not affected. Figure 10 (ii) (c) predicts similar results experimentally.

Figure 10 (i) (d) gives the simulated dynamic performance under 15° phase shift. The estimated frequency shows a peak overshoot of 51.95 Hz at 0.22 s. The overshoot settles down at 0.25 s. Figure 10 (ii) (d) illustrates these results experimentally.

Figure 10 (i) (e) gives the simulated dynamic performance under + 5 Hz frequency change during time interval of 0.2 to 0.4 s. The frequency change leads to a peak oscillation of 55.3 Hz and settles down at 0.25 s. Figure 10 (ii) (e) verifies the results experimentally.

Figure 10 (i) (f) gives the simulated dynamic performance under 10% dc offset during time interval of 0.2 to 0.4 s. The presence of offset is seen in v_s and A_t, where slight distortion in frequency is observed. Phase angle remains unaffected. Figure 10 (ii) (e) verifies these results experimentally.

4.1.1 | Comparative analysis of simulation and experimental results

In this subsection, comparison of the simulated and experimental performance analysis under various grid abnormal conditions for MSRF-PLL and proposed FO-LP MSRF-PLL and FO-LPFO-PI MSRF-PLL is discussed.

Figure 11A performs the comparison of predicted frequencies for three developed PLLs, viz., MSRF-PLL, FO-LP MSRF-PLL, and FO-LPFO-PI MSRF-PLL under 30% voltage sag injected during 0.2 to 0.4 s. MSRF-PLL shows peak overshoot of 51.782 Hz at 0.201 s and settles down at 0.23 s while FO-LPFO-PI MSRF-PLL gives the smallest frequency overshoot of 50.28 Hz and the best performance.

Figure 11B performs the comparison of predicted frequencies under 40% voltage swell during 0.2 to 0.4 s. MSRF-PLL shows the highest frequency overshoot of 53.42 Hz at 0.232 s and settles down at 0.23 s while FO-LPFO-PI MSRF-PLL gives smallest frequency overshoot of 50.6 Hz at 0.203 s and its settling time is 0.22 s.

FO-LP MSRF-PLL has frequency overshoot of 51.8 Hz at 0.204 s settling at 0.25 s. Thus, it is observed that performance of FO-LPFO-PI MSRF-PLL surpasses that of FO-LP MSRF-PLL under swell condition.

Figure 11C predicts the comparison of frequencies under polluted input supply during 0.2 to 0.4 s. MSRF-PLL shows the largest oscillation in the tracked frequency among the three PLLs due to the presence of harmonics. The

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

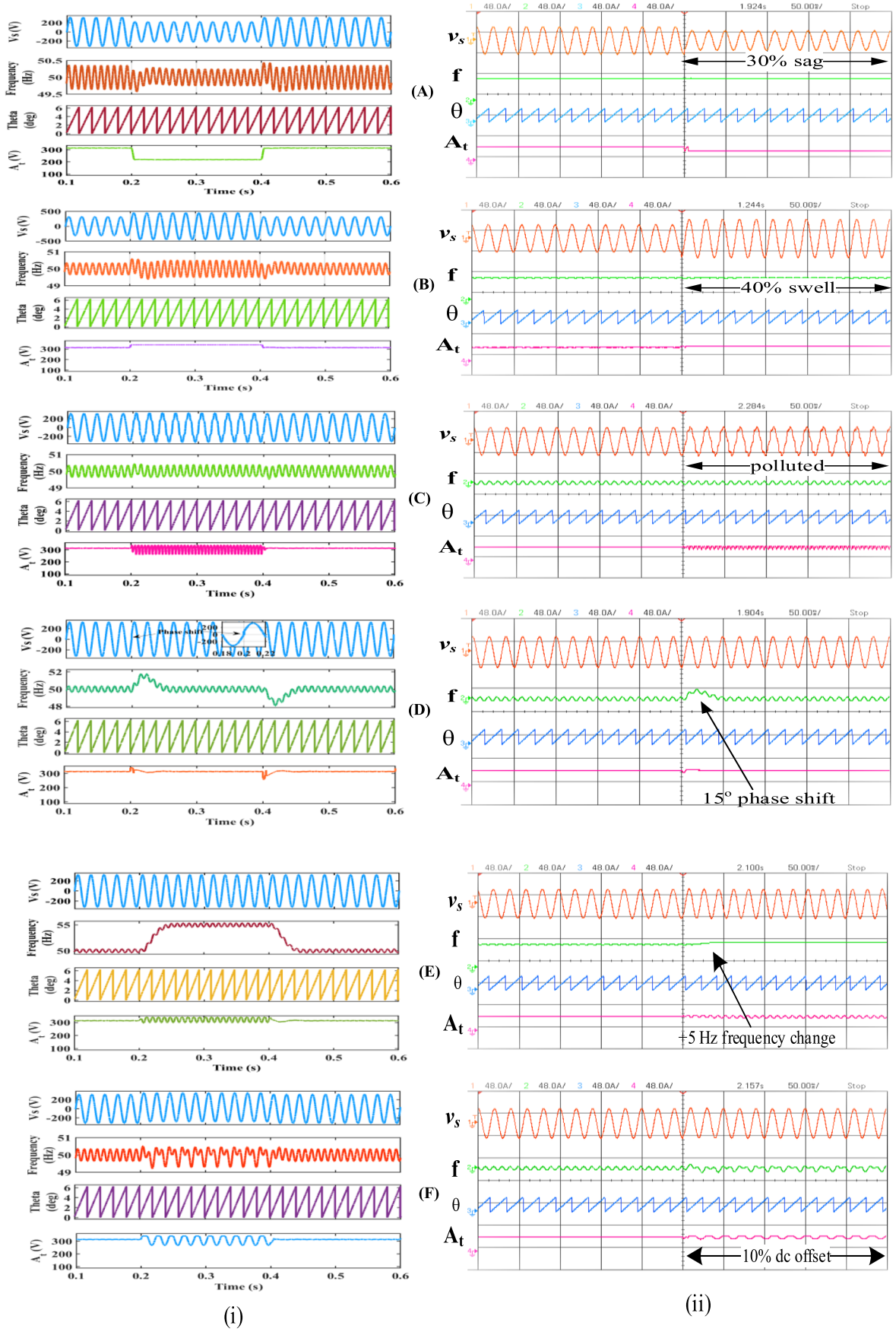


FIGURE 10 Legend on next page.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

FIGURE 10 (i) Simulated and (ii) experimental dynamic performance of FO-LPFO-PI MSRF-PLL under (a) 15% voltage sag, (b) 15% voltage swell, (c) polluted voltage, (d) 15° phase jump, (e) + 5 Hz frequency jump. And (f) 10% dc offset.

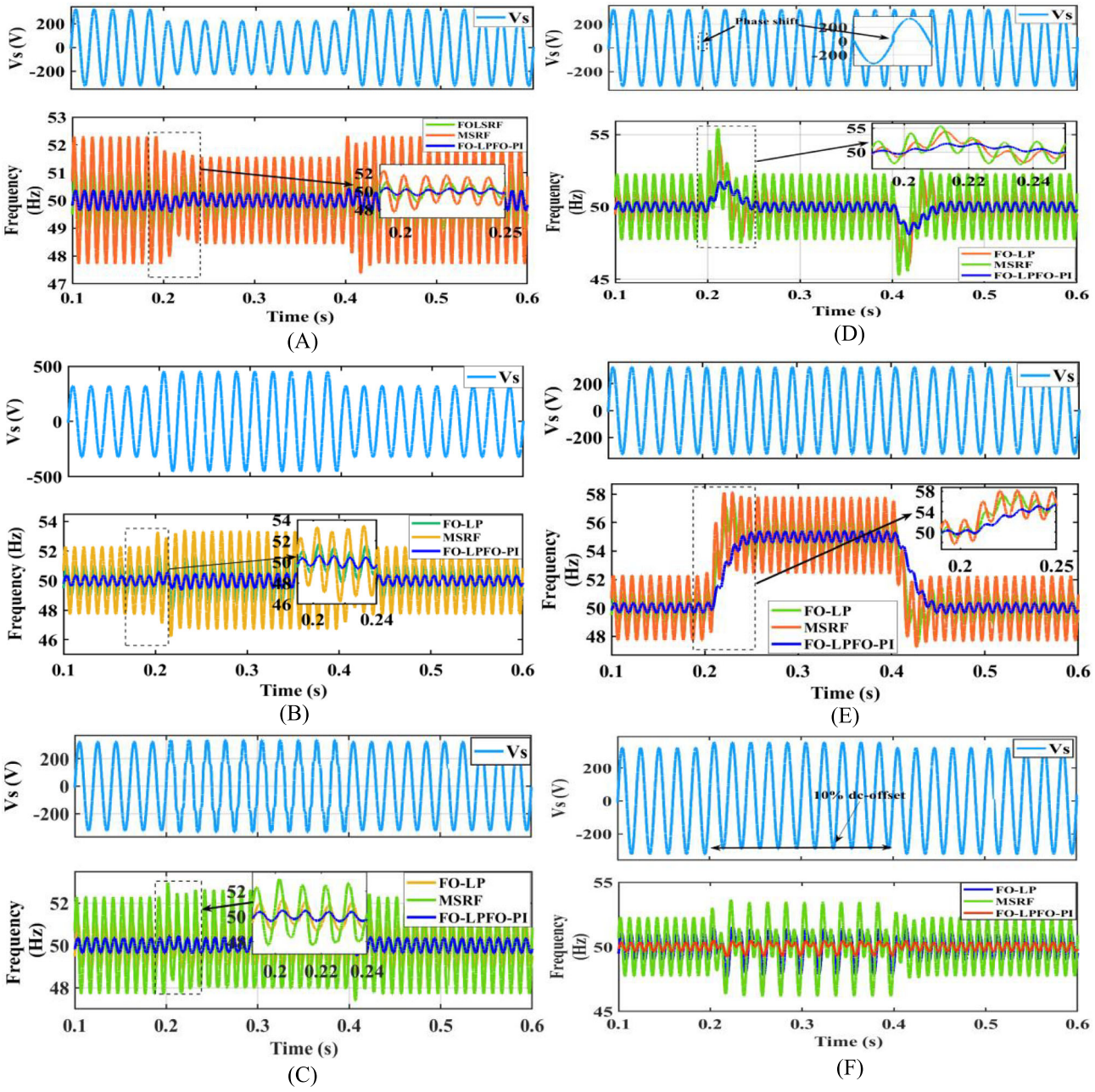


FIGURE 11 Simulated comparative analysis of MSRF-PLL, FO-LPMSRF-PLL, and FO-LPFO-PIMSRF PLL algorithms under (A) 15% voltage sag, (B) 15% voltage swell, (C) figure polluted voltage, (D) 15° phase jumps, (E) + 5 Hz frequency jump, and (F) 10% dc offset.

FO-LPFO-PI MSRF-PLL shows better performance with lower oscillations than FO-LP MSRF-PLL and MSRF-PLL under polluted grid.

Figure 11D predicts the comparison under 15° phase shift in input voltage at 0.2 s. MSRF-PLL shows the highest frequency overshoot of 55.42 Hz at 0.212 s which settles at 0.251 s while FO-LPFO-PI MSRF-PLL gives the smallest frequency overshoot of 51.95 Hz at 0.252 s, and the settling time is 0.22 s. FO-LP MSRF-PLL shows a frequency overshoot of 51.8 Hz at 0.204 s which settles at 0.26 s and shows moderate performance.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 11E shows the comparison under + 5 Hz step frequency shift in input voltage at 0.2 s. MSRF-PLL shows the highest frequency overshoot of 58 Hz at 0.22 s which settles at 0.251 s, while FO-LPFO-PI MSRF-PLL gives the smallest frequency overshoot of 51.95 Hz at 0.252 s, and it settles down at 0.22 s. FO-LP MSRF-PLL shows lowest frequency overshoot of 51.8 Hz at 0.204 s, and it settles down at 0.26 s.

The transient performance is now tested experimentally under various abnormal grid conditions to validate the simulated performance analysis of MSRF-PLL and FO-LP MSRF-PLL. Figure 12 gives the comparison of experimental results under (a) 15% voltage sag, (b) 15% voltage swell, (c) polluted grid voltage, (d) 15° phase jump, (e) + 5 Hz frequency step change, and (f) 10% dc offset. The frequency errors of the integral model of MSRF-PLL, proposed FO-LP MSRF-PLL, and proposed FO-LPFO-PI MSRF-PLL are tracked for comparing performance under grid abnormalities. In all the conditions given in (a) to (f), it is observed that the proposed FO-LPFO-PI MSRF-PLL shows minimum frequency error oscillation which is much lower than that obtained with integer order MSRF-PLL and other proposed fractional FO-LP MSRF-PLL. Hence, the conclusive comparative performance analysis of the three PLLs is shown in Table 5. In this table, the frequency overshoot and settling time for integer order PLL and the proposed FO PLLs are indicated under different grid conditions.

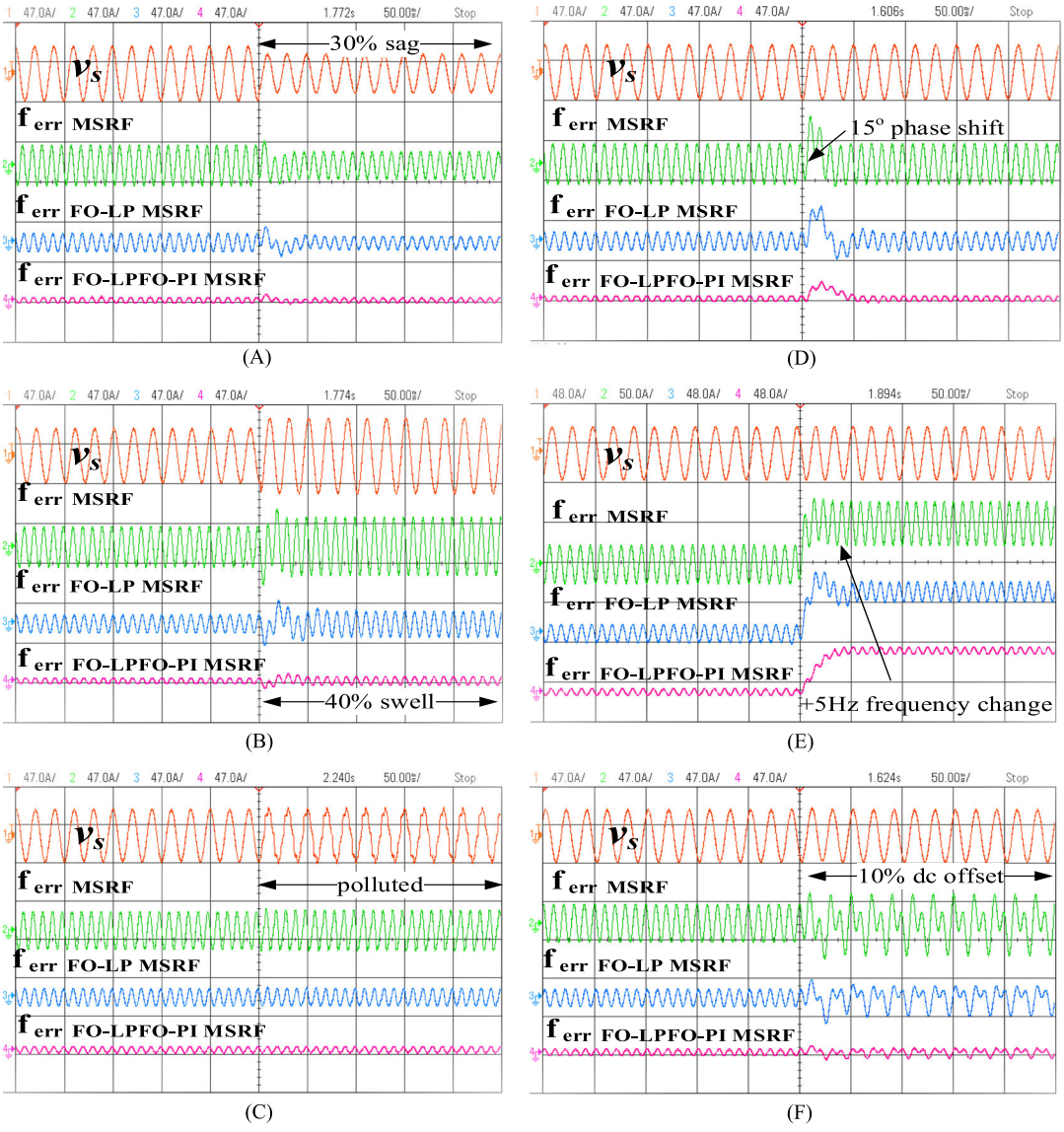


FIGURE 12 Experimental comparative analysis of MSRF-PLL, FO-LP MSRF-PLL, and FO-LPFO-PI MSRF-PLL under various abnormal grid conditions: (A) 15% voltage sag, (B) 15% voltage swell (C) polluted voltage, (D) 15° phase jump, (E) + 5 Hz frequency step change, and (F) 10% dc offset.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

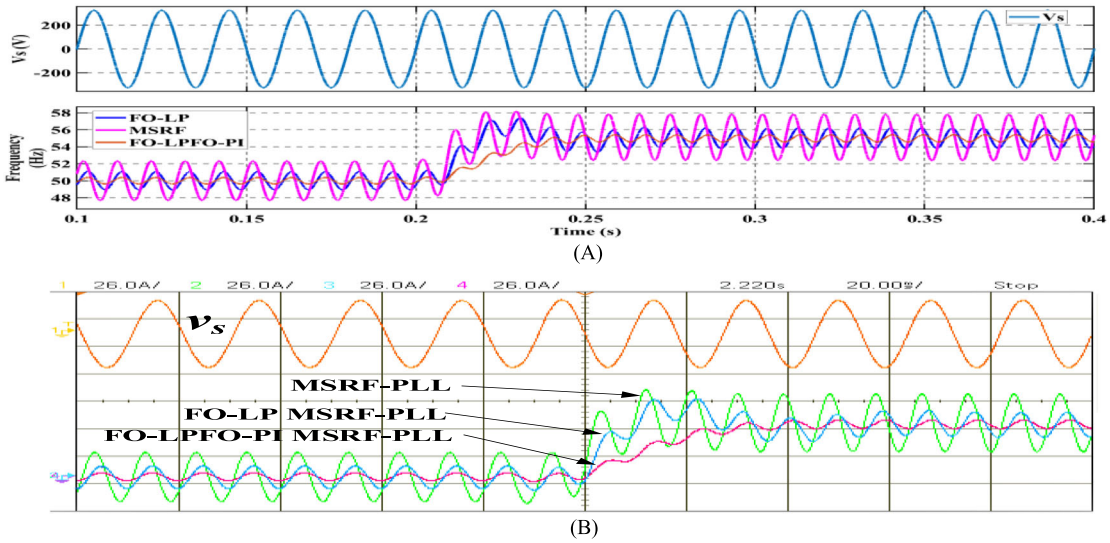


FIGURE 13 (A) Simulated and (B) experimental comparative analysis of MSRF-PLL, FO-LP MSRF-PLL, and FO- LPFO-PI MSRF-PLL under + 5 Hz frequency change.

Figure 13A,B shows the comparative simulation and experimental analysis of MSRF-PLL, FO-LP MSRF-PLL, and FO-LPFO-PI MSRF-PLL under + 5 Hz step change of grid frequency. Hence from Figure 13, it is clearly seen that FO-LP-FO-PI MSRF-PLL gives the best performance among the three PLLs with minimum overshoot and better settling time. The integer-order MSRF-PLL gives the highest overshoot among the three PLLs. The proposed fractional-order PLL (FO-LP MSRF-PLL) has lesser overshoot than MSRF-PLL but it is higher than that of FO-LPFO-PI MSRF-PLL.

5 | CONCLUSION

In the proposed article, the mathematical modeling of three PLLs, viz., integral order MSRF-PLL, fractional FO-LP MSRF-PLL with one FO capacitor a, and FO-LPFO-PI MSRF-PLL with two FO capacitors with both a and b coefficients is discussed. The three PLLs are designed and developed with their open-loop, closed-loop, and error-loop transfer functions. The open-loop and closed-loop Bode diagrams of FO-LP MSRF-PLL and FO-LPFO-PI MSRF-PLL are plotted to predict optimum values of a and b considering PLL stability. The Bode diagrams of FO-LPFO-PI MSRF-PLL are plotted for various values of a at fixed b as 1 along with various values of b for fixed a as 1. The Bode diagrams help to optimize the fractional parameters a and b. The selected optimized values of the real FO coefficients with $a = 1.1$ and $b = 0.1$ are chosen for the proposed FO-LP MSRF-PLL and FO-LPFO-PI MSRF-PLL.

The transient performance analysis for the conventional integral-order MSRF-PLL and proposed FO-LP MSRF-PLL and FO-LPFO-PI is performed under different grid abnormalities like voltage sag and swell, polluted voltage, phase shift, frequency change, and dc offset supply voltage. Both the simulated and experimental comparative analyses are performed rigorously. The comparative results predict that among the two proposed fractional PLLs, the FO-LPFO-PI MSRF-PLL with two FO parameters a and b gives the best performance results during grid abnormalities. Further, this PLL is more stable as compared with FO-LP MSRF-PLL after optimizing the values of a and b parameters.

ACKNOWLEDGEMENT

The authors are thankful to the DST for project SR/FST/ET-I/2022/1026.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Oinam Lotika Devi https://orcid.org/0009-0000-5393-3275

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

1. Bacon VD, Oliveira da Silva SA. Performance improvement of a three-phase phase-locked-loop algorithm under utility voltage disturbances using non-autonomous adaptive filters. *IET Power Electron*. 2015;8(11):2237-2250. doi:10.1049/iet-pel.2014.0808
2. Choi H-J, Song S-H, Jeong S-G, Choi J-Y, Choy I. Design of SRD-F-PLL system with consideration of low-pass filter for high dynamic performance during voltage disturbance. In: *2009 IEEE 6th International Power Electronics and Motion Control Conference*. IEEE; 2009: 720-724. doi:10.1109/IPEMC.2009.5157479
3. Lubura S, Šoja M, Lale S, Ikić M. Single-phase phase locked loop with DC offset and noise rejection for photovoltaic inverters. *IET Power Electron*. 2014;7(9):2288-2299. doi:10.1049/iet-pel.2013.0413
4. Rasheduzzaman M, Khorbotly S, Kimball JW. A modified SRF-PLL for phase and frequency measurement of single-phase systems. In: *2016 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE; 2016:1-7. doi:10.1109/ECCE.2016.7854707
5. Gonzalez-Espin F, Figueres E, Garcera G. An adaptive synchronous-reference-frame phase-locked loop for power quality improvement in a polluted utility grid. *IEEE Trans Ind Electron*. 2012;59(6):2718-2731. doi:10.1109/TIE.2011.2166236
6. Golestan S, Monfared M, Freijedo FD. Design-oriented study of advanced synchronous reference frame phase-locked loops. *IEEE Trans Power Electron*. 2013b;28(2):765-778. doi:10.1109/TPEL.2012.2204276
7. Golestan S, Monfared M, Freijedo FD, Guerrero JM. Advantages and challenges of a type-3 PLL. *IEEE Trans Power Electron*. 2013a; 28(11):4985-4997. doi:10.1109/TPEL.2013.2240317
8. Karimi-Ghartemani M, Karimi H, Khajehoddin SA, Hoseinizadeh SM. Efficient modeling and systematic design of enhanced phase-locked loop structures. *IEEE Trans Power Electron*. 2022;37(8):9061-9072. doi:10.1109/TPEL.2022.3157267
9. Ali Z, Christofides N, Hadjidemetriou L, Kyriakides E, Yang Y, Blaabjerg F. Three-phase phase-locked loop synchronization algorithms for grid-connected renewable energy systems: a review. *Renew Sustain Energy Rev*. 2018;90:434-452. doi:10.1016/j.rser.2018.03.086
10. Guo, X., Wu, W., & Gu, H. (2011). Phase locked loop and synchronization methods for grid- interfaced converters: a review. *PrzeglądElektrotechniczny*, 182–187. https://api.semanticscholar.org/CorpusID:50110456
11. Zou Z-X, Liserre M. Modeling phase-locked loop-based synchronization in grid-interfaced converters. *IEEE Trans Energy Conv*. 2020; 35(1):394-404. doi:10.1109/TEC.2019.2939425
12. Xu J, Qian H, Hu Y, Bian S, Xie S. Overview of SOGI-based single-phase phase-locked loops for grid synchronization under complex grid conditions. *IEEE Access*. 2021;9:39275-39291. doi:10.1109/ACCESS.2021.3063774
13. Chen, Y., Petras, I., &Xue, D. (2009). Fractional order control—a tutorial. 2009 American Control Conference, 1397–1411. 10.1109/ACC.2009.5160719
14. Narayan S, Mehta U, Iro R, Sıkwa'ae H, Kothari K, Singh N. Realization of fractional band pass filter on reconfigurable analog device. *Int Rev Appl Sci Eng*. 2021;13(1):63-69. doi:10.1556/1848.2021.00308
15. El-Khazali R, Ahmad W. Fractional-order phase-locked loop. In: *2007 9th International Symposium on Signal Processing and Its Applications*. IEEE; 2007:1-4. doi:10.1109/ISSPA.2007.4555606
16. Stanisławski R, Latawiec KJ. Normalized finite fractional differences: computational and accuracy breakthroughs. *Int J Appl Math Comput Sci*. 2012;22(4):907-919. doi:10.2478/v10006-012-0067-9
17. Sierociuk D, Wiraszka MS. A new variable fractional-order PI algorithm ∗ ∗this work was supported by the polish National Science Center with the decision number UMO-2014/15/B/ST7/00480. *IFAC-PapersOnLine*. 2018;51(4):745-750. doi:10.1016/j.ifacol.2018.06.207
18. Tseng C-C. Design of variable and adaptive fractional order FIR differentiators. *Signal Process*. 2006;86(10):2554-2566. doi:10.1016/j.sigpro.2006.02.004
19. Pachauri N, Thangavel V, Suresh V, et al. A robust fractional-order control scheme for PV-penetrated grid-connected microgrid. *Mathematics*. 2023;11(6):1283. doi:10.3390/math11061283
20. Narang A, Shah SL, Chen T. Tuning of fractional PI controllers for fractional order system models with and without time delays. In: *Proceedings of the 2010 American Control Conference*. IEEE; 2010:6674-6679. doi:10.1109/ACC.2010.5531353
21. Krishna, B. T. (2012). Fractional calculus based analysis of phase locked loop. 2012 International Conference on Signal Processing and Communications (SPCOM), 1–5. 10.1109/SPCOM.2012.6289990
22. Feng T, Xu C, Yu D. A novel fractional-order charge-pump PLL with the fractional-order loop filter. *Chin Autom Congr (CAC)*. 2020; 2020:3867-3871. doi:10.1109/CAC51589.2020.9326528
23. Pereira LFSC, de Carvalho AA, Perez JL, Batista EA, Godoy RB, de Brito MAG. Design of FOPI controller for the frequency fixed orthogonal signal generator (FFOSG) based PLL through genetic algorithm. In: *2021 14th IEEE International Conference on Industry Applications (INDUSCON)*. IEEE; 2021:1385-1391. doi:10.1109/INDUSCON51756.2021.9529408
24. Mahata S, Kar R, Mandal D. Optimal modelling of (1 + α) order butterworth filter under the CFE framework. *Fractal Fract*. 2020;4(4): 55. doi:10.3390/fractalfract4040055
25. Nako J, Psychalinos C, Elwakil A. 1 + α order generalized butterworth filter structure and its field programmable analog array implementation. *Electronics (Switzerland)*. 2023b;12(5):1225. doi:10.3390/electronics12051225
26. Nako J, Psychalinos C, Elwakil AS, Jurisic D. Design of higher-order fractional filters with fully controllable frequency characteristics. *IEEE Access*. 2023a;11:43205-43215. doi:10.1109/ACCESS.2023.3271863
27. Hassanein AM, Madian AH, Radwan AG, Said LA. On the design flow of the fractional-order analog filters between FPAA implementation and circuit realization. *IEEE Access*. 2023;11:29199-29214. doi:10.1109/ACCESS.2023.3260093
28. Radwan AG, Soliman AM, Elwakil AS. FIRST-order filters generalized to the fractional domain. *J Circuits Syst Comput*. 2008;17(01):55-66. doi:10.1142/S0218126608004162

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

29. Badoni M, Singh A, Pandey S, Singh B. Fractional-order notch filter for grid-connected solar PV system with power quality improvement. *IEEE Trans Ind Electron*. 2022;69(1):429-439. doi:10.1109/TIE.2021.3051585
30. Lazarević, M., Rapaić, M., &Šekara, T. (2014). Introduction to fractional calculus with brief historical background (pp. 3–16).
31. Gorenflo, R., &Mainardi, F. (2008). *Fractional calculus: integral and differential equations of fractional order*.
32. Haubold, H., & Mathai, A. (2017). An introduction to fractional calculus. In *An Introduction to Fractional Calculus*.
33. Ray SS, Atangana A, Noutchie SCO, Kurulay M, Bildik N, Kilicman A. Fractional calculus and its applications in applied mathematics and other sciences. *Math Probl Eng*. 2014;2014:1-2. doi:10.1155/2014/849395

How to cite this article: Devi OL, Singh A. Performance analysis of fractional-order modified SRF PLL under grid abnormalities. *Int J Circ Theor Appl*. 2024;1-21. doi:10.1002/cta.4161

APPENDIX A

Parameters for MSRF-PLL:

Input voltage supply, $v_s = 235$ V, fundamental frequency $= 50$ Hz, $K_p = 0.7$, $K_i = 3$.

Parameters for FO-LP MSRF-PLL:

Input voltage supply, $v_s = 235$ V, fundamental frequency $= 50$ Hz, $\omega_{cf} = 2 \times \pi \times 36.72$, $K_p = 0.7$, $K_i = 0.025$, $a = 1.1$.

Parameters for FO-LPFO-PI MSRF-PLL:

Input voltage supply, $v_s = 235$ V, fundamental frequency $= 50$ Hz, $\omega_{cf} = 2 \times \pi \times 36.72$, $K_p = 0.025$, $K_i = 0.000025$, $a = 1.1$, $b = 0.1$.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4161 by Delhi Technological University, Wiley Online Library on [22/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Water Air Soil Pollut (2024) 235:512
https://doi.org/10.1007/s11270-024-07305-z

Phytoremediation of Chromium (VI)-Contaminated Soil by *Euphorbia tithymaloides* L. and Metagenomic Analysis of Rhizospheric Bacterial Community

Deepika · Anil Kumar Haritash

Received: 20 December 2023 / Accepted: 23 June 2024
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract This study investigated the growth of *Euphorbia tithymaloides* L. in soil contaminated with Cr (VI), focusing on plant development, Cr (VI) accumulation in the plant, and associated rhizosphere bacterial communities. Plants were cultivated in pots for a period of 123 days with varying Cr (VI) levels (0, 10, 20, 30, and 40 mg-kg^{-1}) in the soil. As Cr (VI) concentration increased, plant growth indicators such as fresh weight, dry weight, root length, and shoot length were observed to be decreasing. The Cr (VI) concentration peaked in plant tissues under the 40 mg-kg^{-1} treatment, with concentrations of 397.5 mg-kg^{-1}, 98.12 mg-kg^{-1}, and 62.32 mg-kg^{-1} in roots, stems, and leaves respectively. Whereas, the total Cr (VI) accumulation was maximum in the 30 mg-kg^{-1} treatment due to higher plant biomass. Further, the phytoremediation efficiency of the plant was evaluated through the bio-concentration factor (BCF) and translocation factor (TF). The BCF indicated soil-to-plant heavy metal uptake capacity exceeding 1, while the TF showed root-to-shoot translocation below 1 for all treatments, implying effective phytostabilisation potential. Metagenomic analysis of rhizospheric soil highlighted dominant bacterial phyla of Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, and Acidobacteria. Many observed bacterial taxa are metal-tolerant and commonly found in heavy metal-contaminated soil, suggesting a potential contribution to mitigating metal toxicity and enhancing plant growth. However, a comprehensive investigation into microbial communities' roles during phytoremediation requires further in-depth research.

Keywords Phytoremediation · Chromium (VI) · Soil Contamination · Rhizosphere · Microbial Community · Metagenomic Analysis

Highlights
- The response of *E. tithymaloides* to different levels of Cr (VI) contamination in soil was evaluated through a pot experiment.
- Up to the Cr (VI) concentration of 30 mg-kg^{-1}, *E. tithymaloides* exhibited no significant toxicity to plant health.
- Cr (VI) uptake in plants increased with an increase in concentration.
- BCF was found to be > 1 under all treatments while TF remained < 1, indicating the phytostabilisation potential of the plant.
- Metagenomic analysis of rhizospheric soil highlighted dominant bacterial phyla of Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, and Acidobacteria.

Deepika (✉) · A. K. Haritash
Environmental Microbiology and Bioremediation Laboratory, Department of Environmental Engineering, Delhi Technological University, Bawana Road, Shahbad Daulatpur, Delhi (110042), India
e-mail: deepika_2k20phden502@dtu.ac.in

1 Introduction

Chromium (Cr) is one of the toxic heavy metals which has adverse effects on the environment and health. It enters the environment through various natural and

Published online: 10 July 2024

anthropogenic sources but the release of Cr from man-made activities remains a dominant source because of industrial and mining activities (Wu et al., 2016). Cr has found wide applicability in different industries, including chrome plating, petroleum refining, electroplating, leather tanning, textile manufacturing, wood preservation, production of paints and pigments, etc. (Ramana et al., 2015a, b). These activities inescapably contribute to environmental pollution through the discharge of industrial effluents into water bodies, the lack of management of solid wastes released during the mining process, and the emission of dust into the atmosphere (Karaca et al., 2018). Cr can exist in various oxidation states from 2 to+6 but Cr^{3+} and Cr^{6+} are generally found in the environment (Taneja et al., 2023). The dominantly present Cr (VI) is a strong epithelial irritant and has been categorised as a "group I" or "known" human carcinogen (IARC, 1990), and it is toxic to many plants, aquatic animals, and bacteria (U.S. EPA, 1984). Contrarily, Cr (III) is usually benign and even known to be an essential trace element and cofactor for many enzymes (Ashraf et al., 2019). Both the states are interconvertible in the environment under prevalent conditions, for example, in aerobic conditions Cr exists as Cr (VI) whereas under anaerobic conditions soil organic matter converts Cr (VI) into Cr (III) (Ashraf et al., 2019). Also, Cr (III) is found to be tightly bound with soil particles and forms precipitate with extremely low solubility, which hinders its leaching into groundwater or its absorption by plants, whereas Cr(VI) is highly active and soluble in soil solution (Xia et al., 2019). Therefore, restoration of Cr (VI)-contaminated soils has become very important, and considerable efforts have been made in the past for the development of suitable remediation techniques. Treatment of Cr-contaminated soil could be carried out using various conventional technologies such as soil replacement (Rajendran et al., 2022), immobilization (Vasarevicius et al., 2020), vitrification (ur Rehman et al., 2023), soil washing/flushing (Alaboudi et al., 2020), chemical leaching (Song et al., 2022), electrokinetic remediation (Taneja et al., 2022) and ion-exchange (Qayyum et al., 2020). However, these techniques are prohibitive with respect to time, cost, and production of secondary pollutants (Karaca et al., 2018).

Phytoremediation is an evolving technology that can be considered for the treatment of contaminated soils due to its advantages of being efficient, economical, eco-friendly, aesthetic, and sustainable in nature (Farooq et al., 2020; Zulkernain et al., 2023). This technique uses the remarkable ability of plants to uptake heavy metals from soil (Raj et al., 2020). Phytoremediation stands out as a better option for Cr remediation, particularly when compared to previous technologies. Unlike costly and impractical methods used in the past, phytoremediation offers a cost-effective, scalable solution suitable for larger contaminated areas. Its eco-friendly approach not only minimizes the risk of generating additional pollutants but also requires minimal maintenance and energy inputs once established. Additionally, it does not require the excavation of contaminated sites and offers an *in-situ* remediation. Further, unlike previous physicochemical technologies phytoremediation improves the soil characteristics, fertility, flora and fauna. Among all the processes in phytoremediation, phytoextraction, and phytostabilisation are the most acceptable for the removal of heavy metals from the contaminated soil. The main difference between the two is that phytoextraction uses hyperaccumulator plants to concentrate the contaminants from the soil into their shoots or aboveground biomass (Suman et al., 2018), whereas, in phytostabilisation, the plant reduces the mobility and bioavailability of contaminants (Awa & Hadibarata, 2020). Previously, many plant species have been used for phytoremediation but very few are termed hyperaccumulators as a smaller number of plants exhibited strong tolerance to pollutants along with the ability to transfer contaminants from roots to shoots (Huang et al., 2020). Hyperaccumulators have a 100-fold higher capacity to accumulate heavy metals than generally occurring plants. Additionally, these plants have a translocation factor (TF) greater than 1 (Rezapour et al., 2019). However, the applicability of hyperaccumulators is limited in practice due to their long growth period and low biomass (Farooq et al., 2020). On the other hand, ornamental plants offer an alternative proposition for the treatment of heavy metal-polluted soil due to their rapid growth, higher biomass, high metal tolerance, and non-edible nature (Sehrawat et al., 2021; Zeng et al., 2018). Additionally, these plants may provide economic benefits and improve the aesthetic values of the contaminated site (Mahmood-Ul-Hassan et al., 2020).

Many researchers have studied the feasibility of remediation of Cr-contaminated soils using ornamental plants such as *Polianthes tuberosa* L. (Ramana et al., 2012), *Mirabilis jalapa* L., *Impatiens balsamina* L. and *Tagetes erecta* L. (Miao & Yan, 2013),

Euphorbia mili (Ramana et al., 2015a, b) and *Vinca rosea* L. (Ehsan et al., 2016), etc. Most of the studies report ornamental plants which are seasonal with a short life cycle, therefore, representing lower overall removal efficiency compared to perennially growing ornamental plants. *Euphorbia tithymaloides* L. is an ornamental perennial plant. It is native to tropical and subtropical, North America and Central America, and some areas of South Asia. It is a shrub with zigzag branches and is generally used as an ornamental plant for fencing. It can remove metals from soil throughout the year and since its growth cycle is short, and it can be regularly harvested to export the accumulated 'Cr' out of the phytoremediation system. In addition, the plant is not being used as food for animals or humans, it overcomes the risk of subsequent contamination of food chain. Considering, these environmental benefits, *E. tithymaloides* could be a potential agent for the removal of Cr from contaminated soil.

The rhizospheric bacterial community is another important factor in promoting plant growth and facilitating nutrient availability (Li et al., 2022). The rhizosphere community includes microorganisms like root-associated free living as well as symbiotic rhizobacteria and mycorrhizal fungi. Plant–microbe interaction results in higher microbial diversity and microbial activity in the rhizosphere, even in heavy metal-contaminated soils (Glick, 2020). Previously, many microbes have been used for plant growth promotion and heavy metal immobilization (Chen et al., 2016, 2018). Plants are able to modify their root-associated microbiomes to their advantage even in heavy metal-stressed conditions and contribute significantly to the success of phytoremediation (Manoj et al., 2020). Therefore, the objective of this study was to investigate the phytoremediation of Cr (VI)-contaminated soil by *E. tithymaloides*, and to evaluate the composition of the rhizospheric bacterial community by the 16S rRNA metagenome sequencing method. It is expected that this study will provide information about the phytoremediation potential of *E. tithymaloides* as well as increase our understanding of the *E. tithymaloides*-chromium-soil microbial interactions.

The novelty of the study lies in its exploration of Cr (VI) phytoremediation using ornamental plants, a field that has seen limited attention compared to research focusing on food crops, grasses, and energy crops. Previous studies have primarily overlooked the potential of ornamental plants in remediating Cr-contaminated soil. Specifically, the utilization of *E. tithymaloides* for heavy metal remediation purposes is unprecedented in existing literature. So, this study is the first of its kind to investigate how well *E. tithymaloides* can clean up soil contaminated with Cr (VI).

2 Materials and Methods

2.1 Preparation of Soil and Plant Material

The experiments were carried out in the natural open environment in Delhi, India. The soil samples were collected from a depth of 0–15 cm from the horticulture plots located in the campus of Delhi Technological University, Delhi. Since the site is located in the outskirts of Delhi with minimum/negligible level of contamination of soil by any source. The same was confirmed by analysis of Cr in collected soil samples using atomic absorption spectroscopy (AAS) (Analytik Jena novAA 350 model). After air drying the soil and sieving through the 2 mm sieve, the samples were analyzed for various physicochemical parameters e.g., soil texture (ASTM, 2007), pH, electrical conductivity (EC), total dissolved solids (TDS) (ASTM, 2019), soil organic carbon (SOC) (ASTM, 2014), phosphorus (P) (Bray & Kurtz, 1945), total Kjeldahl nitrogen (TKN) (Blake, 1965), sodium (Na), potassium (K), calcium (Ca), lithium (Li) (APHA, 1992) and Cr (VI) (USEPA, 7196A) (Table 1).

The soil was spiked with a solution of analytical grade $K_2Cr_2O_7$ to prepare five Cr (VI) treatments of 0, 10, 20, 30, and 40 mg-kg^{-1} (T0-T4 treatments). Spiked soil was then allowed to equilibrate for one month with four wetting and drying cycles and intermittent mixing during each cycle. The spiked soil was filled into plastic pots (the bottom diameter: 14.5 cm, top diameter: 19 cm, and height: 17 cm) with 3 kg weight of soil per pot. After that, saplings of *E. tithymaloides* were transplanted into pots keeping three saplings in each pot, and tap water having inborn Cr concentration below the detectable limit (< 0.02 mg/l) was used for watering the plants to avoid any external addition of Cr (VI) during the experiments. In addition, soil spiked with a known concentration (10 mg-kg^{-1}) of Cr (VI) was filled into a pot and was used as a reference sample (without plant). At the end of the experimental period, 95% of Cr (VI) was recovered from the reference sample of soil. The experiment was performed under natural light, no fertilizer was added during the entire period, and tap water was

Table 1 Soil properties before phytoremediation

Soil Properties	Value
Soil Texture	Sandy Loam
pH	7.4
EC (µS/cm)	838
TDS (ppm)	418
SOC (%)	0.4
P (g/kg)	0.21
TKN (%)	0.42
Na (g/kg)	6.69
K (g/kg)	0.23
Ca (g/kg)	0.26
Li (g/kg)	0.0004
Cr (VI) (mg-L^{-1})	BDL* (<0.02)

*Below detectable limit

used for watering the plants @ 200 ml/pot/day. At the end of the analysis, the values of Cr (VI) obtained for soil were corrected against the percent recovery of the reference soil sample.

2.2 Determination of Chlorophyll Content, Growth Parameters, and Cr (VI) Content of Plant

Chlorophyll content in leaves was measured according to Arnon (1949). The weighed fresh leaves were ground well in 80% acetone, and the extract was centrifuged to get a clear solution. A UV–Vis spectrophotometer (LABTRONICS, Model LT-290) was used to determine the absorbance of extracts at wavelengths of 645 and 663 nm. The concentrations of chlorophyll a, chlorophyll b, and total chlorophyll were calculated using the formula suggested by Arnon (1949).

As the plants reached stability in terms of physical growth and completed almost a life-cycle (123 days), the plants were harvested for further investigation. Their roots were removed carefully from the soil and washed several times under running tap water followed by distilled water to remove all the dirt. The root and shoot lengths of all the plants were measured using a metric ruler. The plants were divided into roots, stems, and leaves and oven-dried at 80℃ for 72 h to get a constant weight. The fresh and dry weights of all parts of the plant were also measured and recorded before and after oven drying, respectively.

For extraction of Cr (VI), oven-dried plant samples were ground into a fine powder and thermally oxidised in a muffle furnace at 500℃ temperature for one hour. After that, samples were transferred into the conical flasks and mixed with water in a ratio of 1:10. The mixtures were shaken for two hours in an incubator shaker and then centrifuged at 5000 rpm for 5 min to get a clear extract. Finally, the extracts were analysed for Cr (VI) content in plants by the spectrophotometric method (USEPA, 7196 A). The diphenylcarbazide (DPC) in an acidic medium brings about the reduction of Cr (VI) to Cr3+, and DPC gets oxidised to diphenylcarbazone. The Cr3+-diphenylcarbazone complex formed in this process imparts reddish-violet colour to the solution. The absorbance of the colour solution was measured at 540 nm. The concentration of heavy metal in different plant parts was expressed as mg-kg-1 on a dry weight basis.

2.3 Estimation of Phytoremediation Efficiency

The phytoremediation efficiency of plants was estimated using the BCF, TF, translocation efficiency (TE%), and total heavy metal accumulation rate of plants. The Cr (VI)-tolerance of plant species was evaluated using the tolerance index (TI).

The BCF was described as (Deepika & Haritash, 2023a).

$$BCF = \frac{metal\ concentration\ in\ the\ plant\ (mg/kg)}{metal\ concentration\ in\ the\ soil\ (mg/kg)} \quad (1)$$

The TF showed the ability of plants to translocate heavy metals from the roots to the shoots and was calculated as (Deepika & Haritash, 2023a):

$$TF = \frac{metal\ concentration\ in\ the\ shoot\ (mg/kg)}{metal\ concentration\ in\ the\ root\ (mg/kg)} \quad (2)$$

TE (%) was estimated by using the formula suggested by Meers et al. (2004).

$$TE(\%) = \frac{metal\ content\ in\ the\ shoots\ (mg/kg)}{metal\ content\ in\ the\ whole\ plant\ (mg/kg)} \times 100 \quad (3)$$

The removal of Cr (%) by plant was determined by the formula given (Ramana et al., 2015a, b):

$$\%\ Cr\ removal = \frac{Total\ metal\ uptake\ by\ plant}{Total\ metal\ applied\ to\ the\ soil} \times 100 \quad (4)$$

The total heavy metal accumulation rate (µg g − 1 day^{-1} DW) of plant species can be calculated as follows:

$$Total\ metal\ accumulation\ rate = \frac{DW_s \times C_s + DW_r \times C_r + DW_l \times C_l}{DW_s + DW_r + DW_l \times days\ of\ harvest} \quad (5)$$

Where, DW (g) is dry weight, C ($\mu g/g$) is the concentration of Cr (VI) and the subscript of DW or C denotes the stem (s), root (r) and leaf (l).

TI of plant was calculated by the method adopted by Patra et al. (2020).

$$TI = \frac{Root\ biomass\ of\ treatment\ plants}{Root\ biomass\ of\ control\ plants} \quad (6)$$

2.4 Bacterial Community Analysis by 16S rRNA Metagenome Sequencing

After harvesting the plant, the rhizospheric soil samples were collected from control (T0 treatment) and spiked soil (T4 treatment) to establish the difference in microbial communities. Soil samples were analysed using 16S rRNA (V3-V4) metagenome sequencing method with bioinformatics analysis at Sandor Speciality Diagnostics Pvt. Ltd. Hyderabad, Telangana, India.

Microbial DNA was extracted from the Cr (VI) spiked soil (40 mg-kg^{-1}) and the control soil samples (0 mg-kg^{-1}) using Qiagen Kit Method. Extracted DNA from the samples was subjected to NanoDrop and Gel Check before being taken for PCR amplification. The NanoDrop readings of 260/280 at an approximate value of 1.8 to 2 are used to determine the quality of DNA. The V3-V4 region of 16S rRNA gene was amplified by the primers 16SF (AGAGTTTGATGMTGGCTCAG3') and 16SR (5' TTACCGCGGCMGCSGGCAC3'). The amplification process was carried out using 40 ng of extracted DNA along with 10 pM of each primer by the PCR program. The amplification process was as follows: initial denaturation at 95 ^{0}C, followed by 25 cycles of denaturation at 95 ^{0}C for 15 s, annealing at 60 ^{0}C for 15 s, elongation at 72 ^{0}C for 2 min, final extension at 72 ^{0}C for 10 min and hold at 4 ^{0}C. The raw PCR products were gel-purified and quantified using the Qubit dsDNA high-sensitivity assay kit. Sequencing was performed using Illumina Miseq with 2×300PE v3 sequencing kit.

2.5 Statistical Analysis

All the experiments were performed in triplicates and the data were statistically analysed in MS Excel using the single factor ANOVA and all statistical significance were set at the level of $p < 0.05$. The data presented are means ± standard deviation.

3 Results and Discussion

3.1 Plant Growth

One of the key characteristics of the ideal plant species to be used for the remediation of heavy metal-contaminated soil is its growth performance. In the present study, plant growth has been expressed in terms of biomass (fresh and dry weight) of different plant tissues (Fig. 1) and root and shoot length (Table 2). Higher biomass is a desired parameter for plants being used for phytoremediation, so plants can accumulate more amounts of heavy metals from soil (Cay et al., 2016). Results of this study showed that the application of Cr (VI) negatively affected plant growth compared to the control. The length and biomass of roots and shoots were significantly reduced under the T4 treatment in comparison to the T0 treatment and other treatments in which the concentration of Cr (VI) was low. This implied that up to T3 treatment plant was more adaptable for remediation of Cr (VI)-contaminated soil. The results indicated that Cr (VI) was more toxic to roots than leaves as under T4 treatment, Cr (VI) decreased the dry weight of roots by 92% and leaves by 62.9% over T0 treatment. In contrast to various other heavy metals, Cr exhibits limited mobility within plant roots. As a result, the concentration of Cr within the roots can be up to 100 times greater than in shoots. Within plant structures, Cr (VI) is transformed into Cr (III), which tends to attach to cell walls, preventing the subsequent movement of Cr within the plant's tissues (Sharma et al., 2020). Further, under T1 treatment, root, and shoot length were reduced by 0.91% and 3.1%, respectively, whereas, it was reduced by 37.7% and 41.07% under T4 treatment. Adverse effects of Cr application on plant growth have been previously confirmed by various researchers (Panda et al., 2020; Patra et al., 2020; Ramana et al., 2013, 2015b). Cr (VI) severely affects the membrane (because of high oxidation power), uptake of minerals, and carbon assimilation ability of the plant (Vazquez et al., 1987; Bonet et al., 1991). Prasad et al. (2001) also confirmed that root length was more affected by Cr than other heavy metals and the reason for limited root growth could be the inhibition of root cell division by Cr toxicity.

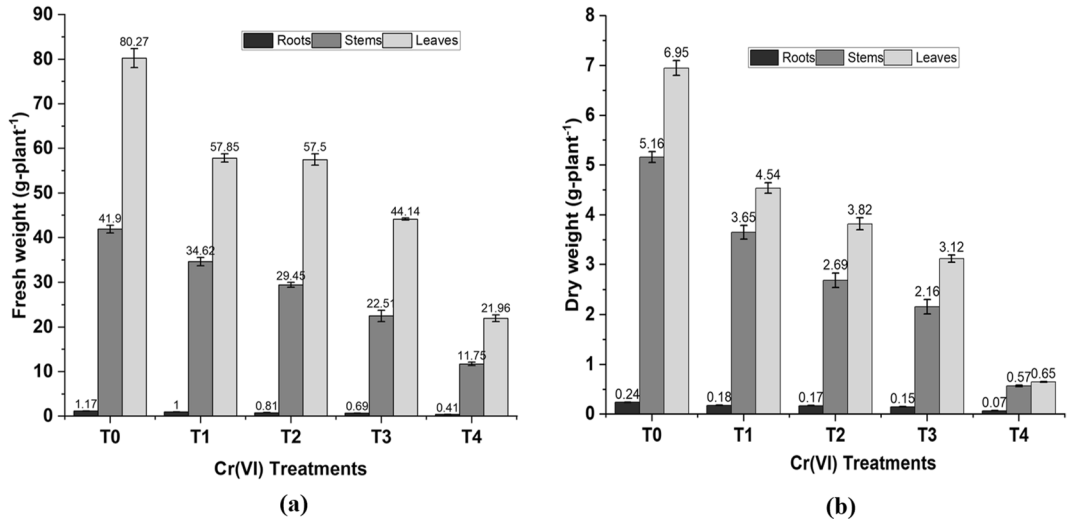


Fig. 1 Effect of Cr (VI) treatment on (**a**) fresh weight and (**b**) dry weight of plant. Mean values are given and error bars indicate standard deviation ($p < 0.05$)

Table 2 Effect of Cr (VI) treatment on root length, shoot length, and chlorophyll content

Soil Treatments	Root Length (cm-plant^{-1})	Shoot length (cm-plant^{-1})	Chl a (mg-g^{-1} FW)	Chl b (mg-g^{-1} FW)	Total Chl (mg-g^{-1} FW)
T0	8.8 ± 1.44	42.0 ± 4	0.99 ± 0.01	0.58 ± 0.05	1.6 ± 0.05
T1	8.8 ± 0.25	40.7 ± 3.21	0.98 ± 0.02	0.42 ± 0.06	1.4 ± 0.07
T2	7.9 ± 0.52	32.9 ± 3.85	0.96 ± 0.01	0.41 ± 0.01	1.4 ± 0.08
T3	7.1 ± 0.05	32.5 ± 2.5	0.95 ± 0.03	0.4 ± 0.02	1.4 ± 0.01
T4	5.5 ± 0.25	24.8 ± 1.15	0.89 ± 0.04	0.36 ± 0.03	1.3 ± 0.08

3.2 Chlorophyll Content

Photosynthesis is an important process for plant growth and biomass production, and the intensity of photosynthesis can be evaluated through the photosynthetic pigment in plant leaves (Liu et al., 2015). It was observed in the present study that a lower level of Cr (VI) did not produce any significant changes in the chlorophyll content of *E. tithymaloides,* but at a higher concentration of 40 mg-kg^{-1} (T4 treatment), the total chlorophyll content of the plant showed 20% reduction compared to T0 treatment (Table 2). A decrease in Chl a, b, and total chlorophyll is a well-documented effect of Cr-stress in plants (Afshan et al., 2015). Cr interferes with the synthesis of delta-aminolevulinic acid (ALA), which is the first step in the biosynthesis of tetrapyrrole leading to the synthesis of heme (Vajpayee et al., 2000). The biosynthesis of chlorophyll might be controlled not only by ALA synthesis, but also by delta-aminolevulinic acid dehydratase (ALAD) activity (Naito et al., 1980). Reduction of chlorophyll content due to altered ALAD activity has been reported in many terrestrial plants (Panda et al., 2020).

3.3 Partitioning of Cr (VI) and Estimation of Phytoremediation Efficiency

While assessing the capacity of plants to treat heavy metal-contaminated soil, it is important to determine the partitioning of heavy metals in the different plant tissues. The Cr (VI) concentration and total accumulation by *E. tithymaloides* are represented in Figs. 2 and 3, respectively. Generally, Cr (VI) concentration

increased in all tissues with an increase in Cr (VI) in soil, and the highest Cr (VI) concentration was found under T4 treatment in all three tissues *i.e.,* roots, stems, and leaves. The results on the partitioning of Cr (VI) showed that roots accumulated higher concentrations of Cr (VI) than stems and leaves indicating the poor translocation of heavy metal within the plant. One of the reasons for a higher concentration of Cr in the roots could be due to the immobilization of Cr in the vacuoles of root cells to render it non-toxic, which may be a natural response of plants (Shankar et al., 2004). Cr (VI) concentration in the roots was around 3 times higher than the stems and leaves together. The Cr (VI) in roots increased consistently from 211 mg-kg^{-1} in T1 treatment to 397.5 mg-kg^{-1} in T4 treatment. Similarly, the concentration of Cr (VI) in stems and leaves also increased uniformly with the increase of Cr (VI) concentration in soil (Fig. 2). These observations were in accord with the previous findings (Ramana et al., 2015a; Subhashini & Swamy, 2013). Lower transport of Cr to aboveground biomass has also been reported in tuberose and some other floriculture plant species such as chrysanthemum, Aster, Dahlia, and Calendula (Ramana et al., 2012, 2013). Further, the total accumulation of Cr (VI) in different tissues of plants was also calculated by multiplying the Cr (VI) concentration in the plant tissue with the respective dry weight and

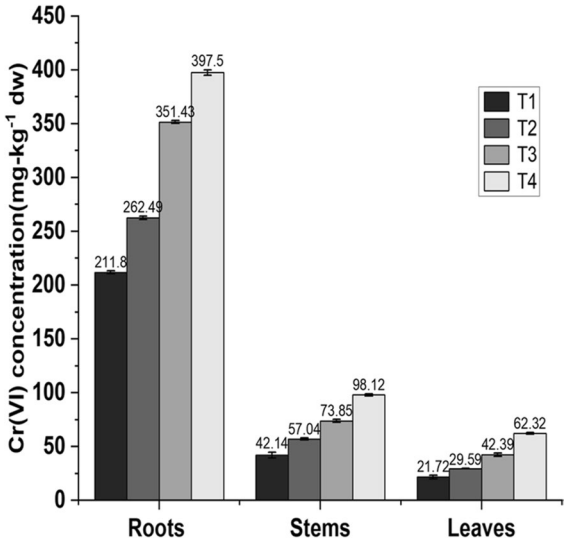


Fig. 2 Cr (VI) concentration in different tissues of *E. tithymaloides*

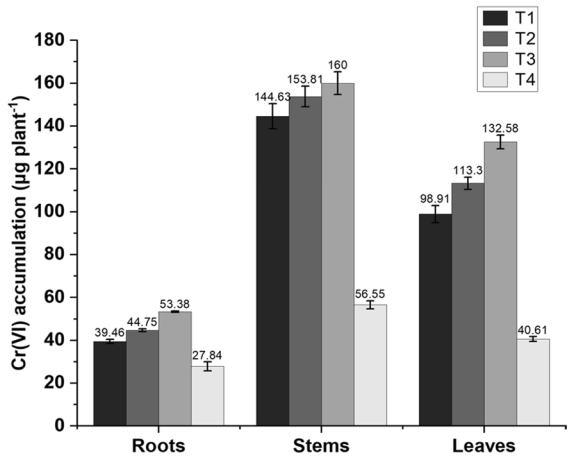


Fig. 3 Total Cr (VI) accumulation by *E. tithymaloides*

expressed as µg-plant^{-1}. The total accumulation of Cr (VI) by plant tissues was lowest under T4 treatment due to the significant reduction in plant biomass. Cr (VI) accumulation was found to be highest under T3 treatment reaching 53.38 µg-plant^{-1}, 160 µg-plant^{-1}, and 132.58 µg-plant^{-1} in roots, stems, and leaves, respectively. These results suggested that the concentration of Cr (VI) under T3 treatment was compliant with the remediation of soil by *E. tithymaloides.*

The phytoremediation efficiency of *E. tithymaloides* was evaluated through BCF, TF, TE, Cr removal (%) by plant, total heavy metal accumulation rate, and TI, which are given below (Table 3). The BCF and the TF are two essential indices for determining if a plant is suitable for phytoextraction or phytostabilization programs. BCF is an index of the ability of a plant to accumulate metal with respect to its concentration in soil and TF is an indication of the ability of a plant to translocate metal from roots to its aboveground tissues. Plants with both a high BCF and a high translocation factor (BCF and TF > 1) can be employed in phytoextraction. Additionally, plants with BCF greater than 1 but TF less than 1 (BCF > 1 and TF < 1) have the potential for phytostabilisation (Deepika & Haritash, 2023b). In the present study, data indicated that Cr accumulated by *E. tithymaloides* was largely retained in roots, as shown by TF < 1, whereas BCF was found to be > 1 across all the treatments. The BCF value was found to be 27.5 for T1 treatment and 13.9 for T4 treatment indicating the reduction in BCF values with an increase in Cr (VI) exposure concentration. Previous studies have also confirmed that BCF decreases with increasing

Table 3 Phytoremediation efficiency measures ($p < 0.05$)

Soil Treatments	BCF $_{(Whole\ plant)}$	TF	TE (%)	% Cr removal by plant	Total metal accumulation rate (μg g − 1 day^{-1} DW)	TI
T1	27.5 ± 0.54	0.30 ± 0.03	23.2 ± 1.3	7.7 ± 0.12	0.28 ± 0.17	67.9 ± 0.73
T2	17.6 ± 0.15	0.33 ± 0.06	24.8 ± 0.3	3.8 ± 0.15	0.37 ± 0.07	54.2 ± 0.6
T3	15.6 ± 0.13	0.33 ± 0.01	24.9 ± 0.61	2.8 ± 0.01	0.51 ± 0.11	44 ± 1
T4	13.9 ± 0.1	0.40 ± 0.03	28.8 ± 0.14	0.6 ± 0.2	0.78 ± 0.12	10.2 ± 0.0

soil heavy metal concentration (Budak et al., 2011; Farooq et al., 2020; Panda et al., 2020). However, the BCF value of *E. tithymaloides* is higher than the other reported studies for ornamental plants grown in Cr-contaminated soil. For example, the highest BCF value of *Euphorbia mili* was found to be 8.46 at 75 mg-kg^{-1} Cr concentration in soil (Ramana et al., 2015a). Similarly, the BCF of *Furcraea gigantea Vent.* was 5.83 for 100 mg-kg^{-1} of Cr in soil (Ramana et al., 2015b).

Further, the TF of *E. tithymaloides* was observed to be < 1, indicating the restricted movement of Cr (VI) to shoots from roots. However, a linear increase (0.3 in T1 treatment to 0.4 in T4 treatment) was observed in TF values in *E. tithymaloides.* Poor translocation of Cr from roots to shoots has been confirmed by other researchers also (Budak et al., 2011; Erenoglu et al., 2007; Patra et al., 2020). In addition, data on TE (%) (Table 3) also showed that the plant was not very efficient in translocating the Cr. Further, the fraction of Cr (VI) removed by *E. tithymaloides* was 7.7% under the T1 treatment and decreased to 0.60% in the T4 treatment. The fraction of Cr (VI) removal under T4 treatment is similar to *Euphorbia mili* (0.6%), a plant of the *Euphorbiaceae* family which has been used for phytoremediation of Cr-contaminated soil (Ramana et al., 2015a). The TI of Cr (VI), which was used to compare the endurance of plants under Cr (VI) stress conditions, gradually decreased with an increase in soil Cr (VI) concentration. Under T1 treatment, TI was 67.89% and then decreased to 54.16%, 44.03%, and 10.22% respectively, under T2, T3, and T4 treatment. A sharp decrease in the TI after T3 treatment implied that the plant could not grow well at higher Cr (VI) concentrations.

3.4 Rhizospheric Bacterial Community

Soil micro-organisms play an important role in the soil–plant system and can affect the phytoremediation of soil contaminants like heavy metals/metalloids by influencing the bioavailability of heavy metals (Nurzhanova et al., 2021). Microbes also help in plant growth and increase plant resistance to heavy metal stress (Tripathi et al., 2020). Compared to bulk soil microbes, rhizosphere microbes are directly involved in providing nutrient elements to plants and are influenced by the release of phytochemicals such as root exudates (Novo et al., 2018). Plants secrete organic compounds that support microbial activity, which tend to provide a beneficial effect on plants (Turner et al., 2013). The enrichment of a beneficial microbial community might be a strategy that enables the plant to survive in toxic environments such as mining-polluted areas (Lin et al., 2021). Therefore, to understand the behaviour of *E. tithymaloides* under a heavy metal-stressed condition, a detailed metagenomic analysis of the rhizosphere bacterial community has been done in the present study.

The 16S rRNA gene sequencing generated 732 valid sequences which could be clustered into 27 phyla, 51 classes, 109 orders, 169 families, and 340 genera. Of these phyla, 5 were considered to be the most dominant in both the soil samples (T0 and T4 treatment), as illustrated in Fig. 4. These phyla included Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, and Acidobacteria, accounting for more than 95% of the classified sequences in the soil samples. In the present study, at the Cr (VI) spiking of 40 mg-kg-1 the order of dominant phyla remained the same. The above-mentioned dominant phyla have been reported in many other studies, however, the rank order might be different. For example, Proteobacteria (27.3%), Chloroflexi (25.9%), Actinobacteria (16.1%), Acidobacteria (13.1%), and Bacteroidetes (7.7%) were the dominant bacterial phyla in the rhizosphere of *Sedum alfredii* Hance grown in multiple heavy metal-contaminated soils (Hou et al., 2017). Similarly, Proteobacteria, Actinobacteria, and Bacteroidetes were found to be the predominant taxa,

when *Brassica juncea* was grown in heavy metal-contaminated mining soil. These taxa were reported to significantly enhance the metal bioavailability and mobilization in soil, therefore, enhancing the metal uptake in plant (Deng et al., 2021; Jeyasundar et al., 2021). Further, the relative proportion and abundance of microbial community is strongly related to the soil properties such as soil texture, pH, organic matter and heavy metal concentration (Xiao et al., 2022). A study by Jiang et al., (2019) confirmed that soils having clay percentage lower than 52% showed high abundance of Firmicutes (54.4%), followed by Proteobacteria (13.2%), and Chloroflexi (8%), whereas when clay percentage were higher than 52%, the dominant bacterial phyla were Actinobacteria (20.3%), Proteobacteria (18.6%), and Firmicutes (16.9%). Proteobacteria, Firmicutes, and Acidobacteria were the most influenced phyla by the soil texture, organic carbon and heavy metal contents in the aforementioned study.

In the present study, Firmicutes was the dominant phylum in both soil samples. The tolerance of Firmicutes to dry environmental conditions might be the reason for its abundance in the soil samples in the present study (Chodak et al., 2015). Also, the lower clay content of the soil (as discussed above) might be another reason for its abundance. Furthermore, the abundance of Firmicutes increased significantly in spiked soil (T4 treatment). These results are in accordance with many previous reports such as on an arsenic-contaminated region (Xiao et al., 2022), an abandoned mine site of rare earth elements (Wei et al., 2019), and an e-waste site in Nigeria (Jiang et al., 2019). Liu et al. (2018) confirmed that Firmicutes are found in abundance in heavy metal-contaminated soil and exhibit strong tolerance to heavy metals. In addition, studies reported that Firmicutes might be important for reducing heavy metal toxicity in soil (Burkhardt et al., 2011). Another phylum that showed a positive correlation with Cr (VI) was Actinobacteria. These results are in agreement with many previous reports which have indicated that Actinobacteria are frequently associated with heavy metal-contaminated environments, most probably due to their high metal tolerance (He et al., 2021; Jiang et al., 2016; Pires et al., 2017). These bacteria are able to promote plant growth by producing bioactive compounds (Alvarez et al., 2017). Some more beneficial features of Actinobacteria and other dominant phyla of the present study are listed in Table 4.

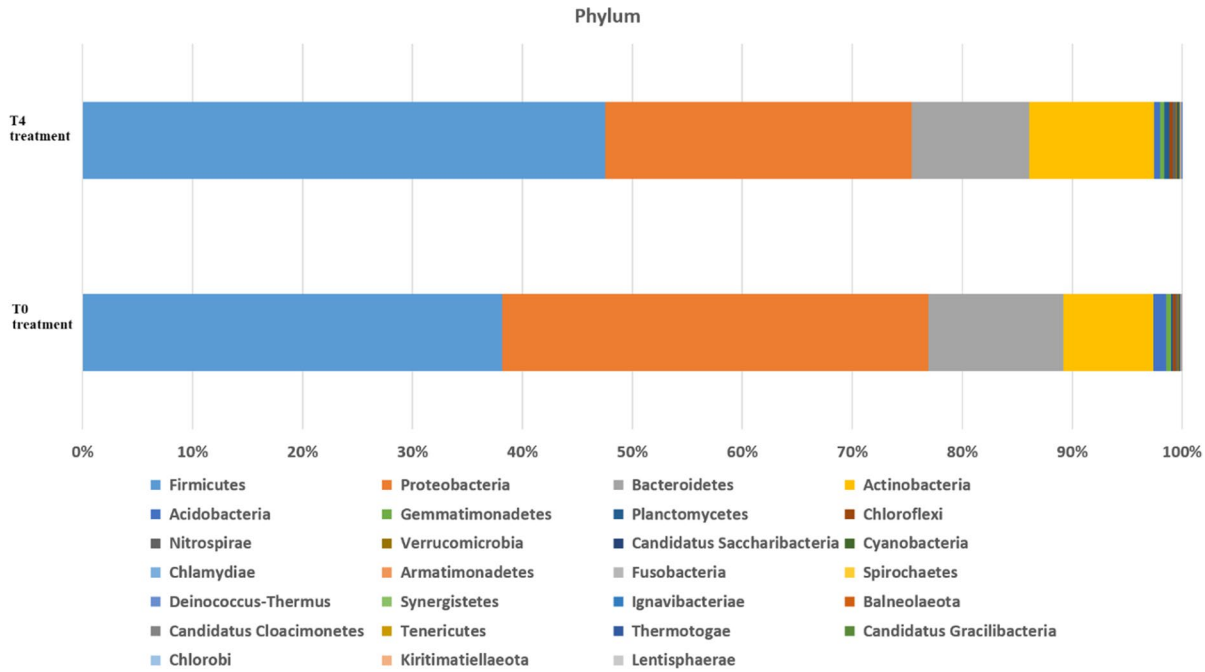


Fig. 4 Relative abundance of dominant bacterial phyla in T4 and T0 soil treatments

Phylum Proteobacteria is reported to be one of the most resistant phyla to heavy metal contamination such as cadmium (Cd), zinc (Zn), arsenic (As), copper (Cu), lead (Pb), and can be found in many mining areas (Chen et al., 2018; Ma et al., 2015; Narendrula-Kotha & Nkongolo, 2017; Yan et al., 2016). Likewise, Acidobacteria has been reported to be predominant in heavy metal-contaminated soil (Ward et al., 2009; Briceño et al., 2012). These are acidophilic in nature and ubiquitous in various ecosystems although these are reported to be abundant in high-nutrient soil (Wei et al., 2019). In the present study, it has been reported that Firmicutes, Proteobacteria, Bacteroidetes, and Acidobacteria are found relatively more (in the category of top 10 phylum); but Proteobacteria and Acidobacteria were slightly more in percentage compared to the spiked soil (40 mg-kg^{-1}), although the difference is very little. The difference may be ascribed to several factors (edaphic, biological, and chemical) which regulate bacterial diversity. Since there is no toxicity of Cr (VI) in control, a slightly higher fraction of Acidobacteria and Proteobacteria might be observed. Further, Bacteroidetes had no significant difference in both soil samples. Bacteroidetes have been found in heavy metal-polluted soil (Haller et al., 2011; Ma et al., 2015), and due to their presence at metal-contaminated and uncontaminated sites, researchers indicated that this phylum may possess metal-resistant properties (Ni et al., 2016). Wang et al. (2018) reported that Proteobacteria, Firmicutes, and Bacteroidetes are the important bacterial phyla that help in decreasing the toxicity of heavy metals.

At the genus level, a total of 340 genera were obtained in both the soil samples, and 228 genera were common to both soil samples, accounting for 67.1% of the total classified genera. The exclusive genera for soils with T4 treatment and T0 treatment were only 67 and 45, respectively. As shown in Fig. 5, the dominant genera were selected from each soil sample (accounting for approximately 80% of total classified genera) and their abundance was compared with each other. Among the dominant genera, the relative abundance of *Clostridium, Clostridioides, Bifidobacterium, Lactobacillus, Prevotella, Streptococcus, Megasphaera, Stenotrophomonas, Faecalitalea, Collinsella, Faecalibacterium, Weissella, Bacillus, pediococcus* and *Olsenella* was higher in soils with T4 treatment as compared to T0 treatment. However, the abundance of *Acinetobacter, Flavobacterium, Burkholderia, Sphingomonas,* and *Pseudomonas* was lower in the T4 treatment than T0 treatment. Additionally, *Janthinobacterium* was exclusively present in soil with T0 treatment. Many of

Table 4 Some beneficial features of dominant bacterial phyla of the present study

Phylum	Beneficial features
Firmicutes	-Weathering of minerals and nutrient cycling (Uroz et al., 2011) -Abundant in low-nutrient soil (Wei et al., 2019) -Participate in the degradation of macromolecular organic matter (Fan et al., 2016) -Important for ameliorating heavy metal toxicity (Burkhardt et al., 2011; Wang et al., 2018) -Inhibits soil-borne disease (Liu et al., 2018) -Resistant to drought and can survive in extreme environments (Chodak et al., 2015)
Proteobacteria	-Affect soil carbon and nitrogen cycling as well as enzymatic activities (Acosta-Martinez et al., 2010) -Most metal-tolerant and dominant bacterial groups in the mine tailings (Hamamura et al., 2013; Wei et al., 2019)
Bacteroidetes	-Indicator of soil health -Change physical and chemical characteristics of soil including soil pH and organic components in the rhizosphere (Gao et al., 2021; Wang et al., 2018) -Helps in the degradation of high-molecular-weight organic matter in the rhizosphere (Naumof & Dedysh, 2012)
Actinobacteria	-Influence pathogenic suppression and degradation of recalcitrant compounds - Heavy metal resistant due to their metabolic versatility and rapid colonization ability (Jeyasundar et al., 2021) -Promote plant growth by producing bioactive compounds (Wang et al., 2021) -Producer of many antibiotics and may enhance the plant immune system (Chen et al., 2018)
Acidobacteria	-Helps in the excretion of extracellular polymeric substances (EPS-polysaccharides and proteins) to detoxify Cr (Zhang et al., 2021) -Promotes plant hormone (indole-3-acetic acid) production and facilitates iron acquisition in the plant (Gong et al., 2021) -Disintegrators of organic matter and provides resilience in nutrient-deprived environments (Rawat et al., 2012)

the above-mentioned genus are reported to be associated with heavy metal-contaminated soil. For example, *Acinetobacter* have been frequently found in metal-polluted ecosystems (Brantner & Senko, 2014). Many strains of this genus are involved in mineral weathering, phosphate solubilization, iron chelation (Uroz et al., 2011) and degradation of contaminants (Simonsen et al., 2012). Many reports have highlighted the role of various bacterial genera in the phytoremediation of heavy metal-contaminated soil by enhancing the plant growth and metal uptake by plant. Previous reports on enhanced phytoremediation by some of the dominant genera of this study has been given in Table 5. Although, the role of some genera such as *Clostridium, Clostridioides, Bifidobacterium, Prevotella,* and *Blautia* is still unknown in the process of phytoremediation.

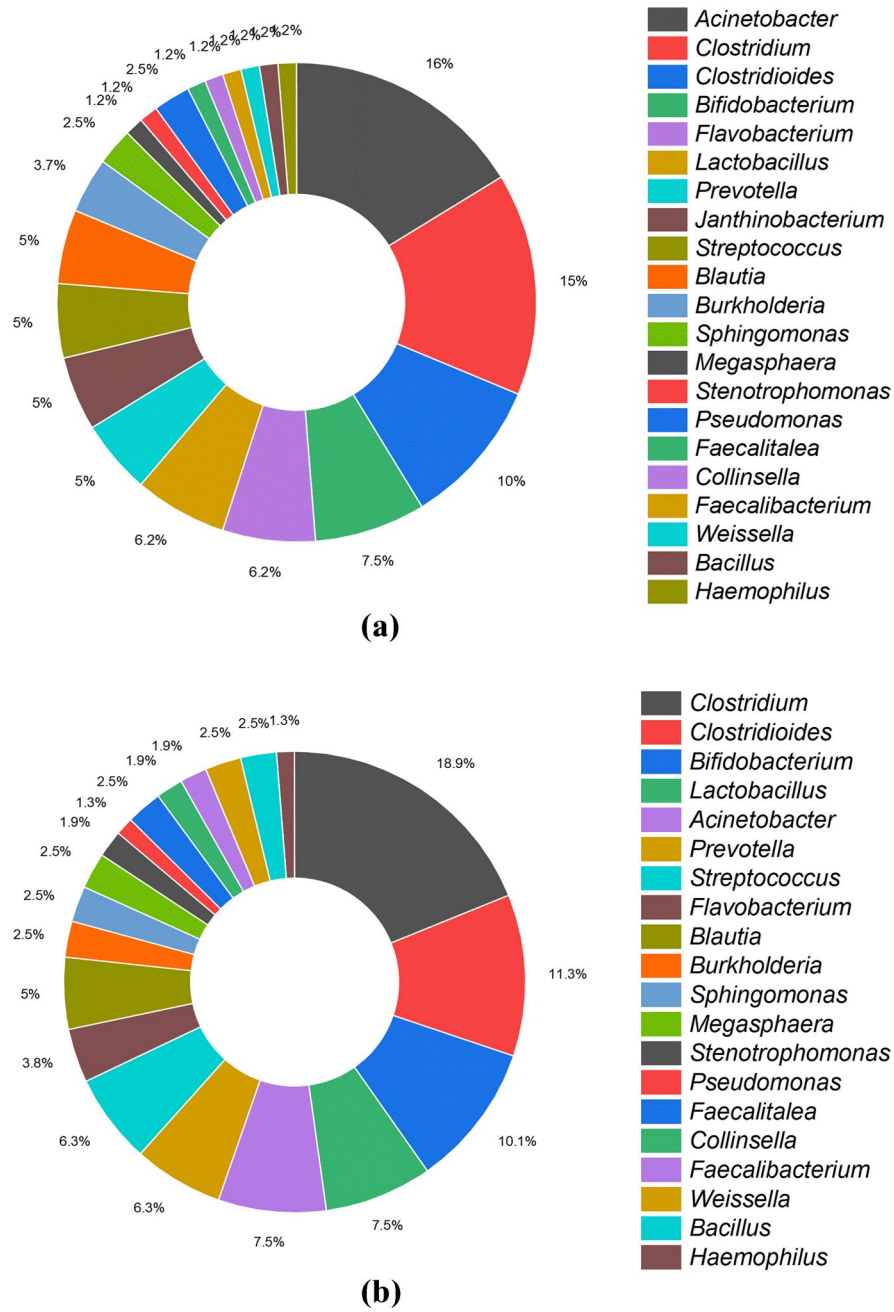


Fig. 5 Dominant genera of soils in (**a**) T0 treatment (control) and (**b**) T4 treatment (40 mg-kg-1)

Table 5 Previous reports on enhanced phytoremediation by some dominant genera of the present study

Genera	Associated plant (s)	Heavy Metal (s)	Beneficial features	Reference
Acinetobactor	*Catharanthus longifolius*	Cd	-Phosphate solubilization, ACCD activity, auxin, and siderophores production -Improved chlorophyll, soluble proteins, and sugar content -Higher metal uptake and translocation	(Yasin et al., 2019)
Flavobacterium	*Orychophragmus violaceus*	Zn	-Stimulated root elongation -Increased mobility and enhanced accumulation of Zn	(He et al., 2010)
Lactobacillus	*Brassica juncea* L	Cd & Zn	-decreased soil pH -enhanced plant growth and enzyme activity -enhanced phytoavailability of Cd and Zn	(Zhang et al., 2023)
Janthinobacterium	*Salix caprea* L	Zn, Pb & Cd	-IAA (indole acetic acid) production -showed resistant to high heavy metal concentration	(Kuffner et al., 2008)
Streptococcus	*Beta vulgaris* L	Cd, Zn & Cu	-Enhanced heavy metal tolerance and accumulation -Overproduction of GSH and phytochelatin	(Liu et al., 2015)
Burkholderia	*Catharanthus roseus* L	Cd	-Enhanced plant growth -Reduced production of MDA and H_2O_2 -ACCD activity, phosphate solubilization, auxin, and siderophore production -Enhanced the amount of water extractable Cd from soil	(Khan et al., 2018)

4 Conclusion

In conclusion, the findings of this study highlight the response of *E. tithymaloides* to different levels of Cr (VI) contamination. The results indicate that at concentration up to 30 mg-kg-1 (T3 treatment), *E. tithymaloides* exhibited significant growth, suggesting its adaptability for the remediation of Cr (VI)-contaminated soils. Phytoremediation efficiency assessment revealed that *E. tithymaloides* is not suited for phytoextraction, as it could not meet the criteria of TF value being more than 1. However, the observed BCF value exceeding 1 suggests that the plant is well-suited for phytostabilization. Further, metagenomic analysis of the rhizosphere revealed the prevalence of heavy metal-tolerant bacterial phyla, such as Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, and Acidobacteria, in both T4 and T0 treatment soil samples. These bacterial taxa are recognized for their tolerance to heavy metals and their frequent association with metal-contaminated soils. This abundance of metal-tolerant bacterial phyla in the rhizosphere likely contributes to enhancing E. tithymaloides' tolerance under Cr (VI) stress.

Taken together, the outcomes of this study establish *E. tithymaloides* as a promising perennial ornamental plant suitable for phytostabilization of Cr (VI) contaminated soil ($\leq$30 mg-kg-1). Although several studies have reported efficient removal of Cr (VI) for concentrations up to 200 mg-kg^{-1} in contaminated soil using different species of ornamental plants, *E. tithymaloides* will be effective in Phytostabilisation of Cr (VI) up to a concentration of 30 mg-kg^{-1} without any effect on plant health. The coexistence of metal-tolerant bacterial phyla within the rhizosphere further underscores the plant's role in ameliorating metal toxicity in polluted soils. This research highlights the potential of using *E. tithymaloides* in ecological restoration efforts for Cr (VI)-contaminated environments, with implications for sustainable remediation strategies and environmental health.

Acknowledgements We are thankful to Delhi Technological University, Delhi, India for providing facilities to carry out the research.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability All data used or analysed during this study are included in this article and its supplementary information files.

Declarations

Ethics Approval Not applicable.

Consent to Publish Not applicable.

Consent to Participate Not applicable.

Competing Interest The authors have no relevant financial or non-financial interests to disclose.

References

Acosta-Martínez, V., Dowd, S. E., Bell, C. W., Lascano, R., Booker, J. D., Zobeck, T. M., & Upchurch, D. R. (2010). Microbial community composition as affected by dryland cropping systems and tillage in a semiarid sandy soil. *Diversity, 2*(6), 910–931. https://doi.org/10.3390/d2060910

Afshan, S., Ali, S., Bharwana, S. A., Rizwan, M., Farid, M., Abbas, F., & Abbasi, G. H. (2015). Citric acid enhances the phytoextraction of chromium, plant growth, and photosynthesis by alleviating the oxidative damages in Brassica napus L. *Environmental Science and Pollution Research, 22*(15), 11679–11689. https://doi.org/10.1007/s11356-015-4396-8

Alaboudi, K. A., Ahmed, B., & Brodie, G. (2020). Soil washing technology for removing heavy metals from a contaminated soil: A case study. *Journal of Environmental Studies, 29*(2), 1029–1036. https://doi.org/10.15244/pjoes/104655

Alvarez, A., Saez, J. M., Costa, J. S. D., Colin, V. L., Fuentes, M. S., Cuozzo, S. A., & Amoroso, M. J. (2017). Actinobacteria: Current research and perspectives for bioremediation of pesticides and heavy metals. *Chemosphere, 166*, 41–62. https://doi.org/10.1016/j.chemosphere.2016.09.070

APHA. (1992). *Method 3111. Standard methods for the examination of water and wastewater* (19th ed.). American Public Health Association.

Arnon, D. I. (1949). Copper enzymes in isolated chloroplasts. Polyphenoloxidase in Beta vulgaris. *Plant Physiology, 24*(1), 1. https://doi.org/10.1104/pp.24.1.1

Ashraf, S., Ali, Q., Zahir, Z. A., & Asghar, H. N. (2019). Phytoremediation: Environmentally sustainable way for reclamation of heavy metal polluted soil. *Ecotoxicology and Environmental Safety, 174*, 714–727. https://doi.org/10.1016/j.ecoenv.2019.02.068

ASTM. (2007). D422-63: Standard test method for particle-size analysis of soils. Annual book of ASTM standards, 4. https://doi.org/10.1520/D0422-63R07E01

ASTM. (2014). D 2974–14: Standard test methods for moisture, ash, and organic matter of peat and other organic soils. *ASTM Int,* 1–4. https://doi.org/10.1520/D2974-14

ASTM. (2019). D4972-19: Standard test methods for pH of soils. *ASTM Int,* 1–6. https://doi.org/10.1520/D4972-19

Awa, S. H., & Hadibarata, T. (2020). Removal of heavy metals in contaminated soil by phytoremediation mechanism: a review. *Water, Air, & Soil Pollut, 231*, 47. https://doi.org/10.1007/s11270-020-4426-0

Blake, G. R. (1965). Bulk density. *Methods of soil analysis: Part 1 physical and mineralogical properties, including statistics of measurement and sampling, 9*, 374–390. https://doi.org/10.2134/agronmonogr9.1.c30

Bonet, A., Poschenrieder, C., & Barcelo, J. (1991). Chromium III-iron interaction in Fe-deficient and Fe-sufficient bean plants. I. Growth and nutrient content. *Journal of Plant Nutrition, 14*(4), 403–414. https://doi.org/10.1080/01904169109364211

Brantner, J. S., & Senko, J. M. (2014). Response of soil-associated microbial communities to intrusion of coal mine-derived acid mine drainage. *Environmental Science & Technology, 48*(15), 8556–8563. https://doi.org/10.1021/es502261u

Bray, R. H., & Kurtz, L. T. (1945). Determination of total, organic, and available forms of phosphorus in soil. *Soil Science, 59*, 39–45.

Briceño, G., Fuentes, M. S., Palma, G., Jorquera, M. A., Amoroso, M. J., & Diez, M. C. (2012). Chlorpyrifos biodegradation and 3, 5, 6-trichloro-2-pyridinol production by actinobacteria isolated from soil. *International Biodeterioration & Biodegradation, 73*, 1–7. https://doi.org/10.1016/j.ibiod.2012.06.002

Budak, F., Zaimoglu, Z., & Basci, N. (2011). Uptake and translocation of hexavalent chromium by selected species of ornamental plants. *Polish Journal of Environmental Studies, 20*(4), 857–862.

Burkhardt, E. M., Bischoff, S., Akob, D. M., Büchel, G., & Küsel, K. (2011). Heavy metal tolerance of Fe (III)-reducing microbial communities in contaminated creek bank soils. *Applied and Environmental Microbiology, 77*(9), 3132–3136. https://doi.org/10.1128/AEM.02085-10

Cay, S., Uyanik, A., & Engin, M. (2016). EDTA supported phytoextraction of Cd from contaminated soil by four different ornamental plant species. *Soil and Sediment Contamination: An International Journal*. https://doi.org/10.1080/15322383.2016.1138448

Chen, L., Brookes, P. C., Xu, J., Zhang, J., Zhang, C., Zhou, X., & Luo, Y. (2016). Structural and functional differentiation of the root-associated bacterial microbiomes of perennial ryegrass. *Soil Biology and Biochemistry, 98*, 1–10. https://doi.org/10.1016/j.soilbio.2016.04.004

Chen, Y., Ding, Q., Chao, Y., Wei, X., Wang, S., & Qiu, R. (2018). Structural development and assembly patterns of the root-associated microbiomes during phytoremediation. *Science of the Total Environment, 644*, 1591–1601. https://doi.org/10.1016/j.scitotenv.2018.07.095

Chodak, M., Gołębiewski, M., Morawska-Płoskonka, J., Kuduk, K., & Niklińska, M. (2015). Soil chemical properties affect the reaction of forest soil bacteria to drought and rewetting stress. *Annals of Microbiology, 65*(3), 1627–1637. https://doi.org/10.1007/s13213-014-1002-0

Deepika, & Haritash, A. K. (2023a). Cadmium uptake from soil by ornamental metallophytes: a meta-analytical approach. *Environmental Management, 71*, 1087–1097. https://doi.org/10.1007/s00267-022-01776-8

Deepika, & Haritash, A. K. (2023b). Phytoremediation potential of ornamental plants for heavy metal removal from contaminated soil: a critical review. *Horticulture, Environment, and Biotechnology*. https://doi.org/10.1007/s13580-023-00518-x

Deng, R., Huang, D., Lei, L., Zhou, C., Yin, L., Liu, X., Chen, S., Li, R., & Tao, J. (2021). Stabilization of lead in polluted sediment based on an eco-friendly amendment strategy: Microenvironment response mechanism. *Journal of Hazardous Materials, 415*, 125534. https://doi.org/10.1016/j.jhazmat.2021.125534

Ehsan, N., Nawaz, R., Ahmad, S., Khan, M. M., & Hayat, J. (2016). Phytoremediation of chromium-contaminated soil by an ornamental plant, Vinca (*Vinca rosea L.*). *Journal of Environmental and Agricultural Sciences, 7*, 29–34.

Erenoglu, B. E., Patra, H. K., Khodr, H., Römheld, V., & Von Wiren, N. (2007). Uptake and apoplastic retention of EDTA-and phytosiderophore-chelated chromium (III) in maize. *Journal of Plant Nutrition and Soil Science, 170*(6), 788–795. https://doi.org/10.1002/jpln.200700068

Fan, M., Lin, Y., Huo, H., Liu, Y., Zhao, L., Wang, E., & Wei, G. (2016). Microbial communities in riparian soils of a settling pond for mine drainage treatment. *Water Research, 96*, 198–207. https://doi.org/10.1016/j.watres.2016.03.061

Farooq, A., Nadeem, M., Abbas, G., Shabbir, A., Khalid, M., Javeed, H. M., & Akhtar, G. (2020). Cadmium partitioning, physiological and oxidative stress responses in marigold (*Calendula calypso*) grown on contaminated soil: Implications on phytoremediation. *Bulletin of Environmental Contamination and Toxicology, 105*, 270–276.

Gao, T. P., Wan, Z. D., Liu, X. X., Fu, J. W., Chang, G. H., Sun, H. L., & Fang, X. W. (2021). Effects of heavy metals on bacterial community structure in the rhizosphere of Salsola collina and bulk soil in the Jinchuan mining area. *Geomicrobiology Journal, 38*(7), 620–630. https://doi.org/10.1080/01490451.2021.1914784

Glick, B. R. (2020). Introduction to plant growth-promoting bacteria. *Beneficial plant-bacterial interactions* (pp. 1–37). Cham: Springer. https://doi.org/10.1007/978-3-030-44368-9_1

Gong, X., Huang, D., Liu, Y., Zou, D., Hu, X., Zhou, L., & Xiao, Z. (2021). Nanoscale zerovalent iron, carbon nanotubes and biochar facilitated the phytoremediation of cadmium contaminated sediments by changing cadmium fractions, sediments properties and bacterial community structure. *Ecotoxicology and Environmental Safety, 208*, 111510. https://doi.org/10.1016/j.ecoenv.2020.111510

Haller, L., Tonolla, M., Zopfi, J., Peduzzi, R., Wildi, W., & Poté, J. (2011). Composition of bacterial and archaeal communities in freshwater sediments with different contamination levels (Lake Geneva, Switzerland). *Water Research, 45*(3), 1213–1228. https://doi.org/10.1016/j.watres.2010.11.018

Hamamura, N., Fukushima, K., & Itai, T. (2013). Identifcation of Antimony- and Arsenic-Oxidizing Bacteria Associated with Antimony Mine Tailing. *Microbes and Environments, 28*, 257–263.

He, C. Q., Tan, G. E., Liang, X. D. U. W., Du, W., Chen, Y. L., Zhi, G. Y., & Zhu, Y. (2010). Effect of Zn-tolerant bacterial strains on growth and Zn accumulation in Orychophragmus violaceus. *Applied Soil Ecology, 44*(1), 1–5.

He, L., Zhu, Q., Wang, Y., Chen, C., He, M., & Tan, F. (2021). Irrigating digestate to improve cadmium phytoremediation potential of Pennisetum hybridum. *Chemosphere, 279*, 130592. https://doi.org/10.1016/j.chemosphere.2021.130592

Hou, D., Wang, K., Liu, T., Wang, H., Lin, Z., Qian, J., & Tian, S. (2017). Unique rhizosphere micro-characteristics facilitate phytoextraction of multiple metals in soil by the hyperaccumulating plant Sedum alfredii. *Environmental Science & Technology, 51*(10), 5675–5684. https://doi.org/10.1021/acs.est.6b06531

Huang, L., Wang, Q., Zhou, Q., Ma, L., Wu, Y., Liu, Q., & Feng, Y. (2020). Cadmium uptake from soil and transport by leafy vegetables: A meta-analysis. *Environmental Pollution, 264*, 114677. https://doi.org/10.1016/j.envpol.2020.114677

International Agency for Research on Cancer (IARC). (1990). Chromium, nickel and welding. *IARC monographs on the evaluation of carcinogenic risks to humans.* (Vol. 49). Lyon

Jeyasundar, P. G. S. A., Ali, A., Azeem, M., Li, Y., Guo, D., Sikdar, A., & Zhang, Z. (2021). Green remediation of toxic metals contaminated mining soil using bacterial consortium and Brassica juncea. *Environmental Pollution, 277*, 116789. https://doi.org/10.1016/j.envpol.2021.116789

Jiang, L., Song, M., Yang, L., Zhang, D., Sun, Y., Shen, Z., & Zhang, G. (2016). Exploring the Influence of Environmental Factors on Bacterial Communities within the Rhizosphere of the Cu-tolerant plant, Elsholtzia splendens. *Scientific Reports, 6*(1), 1–10. https://doi.org/10.1038/srep36302

Jiang, B., Adebayo, A., Jia, J., Xing, Y., Deng, S., Guo, L., & Zhang, D. (2019). Impacts of heavy metals and soil properties at a Nigerian e-waste site on soil microbial community. *Journal of Hazardous Materials, 362*, 187–195. https://doi.org/10.1016/j.jhazmat.2018.08.060

Karaca, O., Cameselle, C., Reddy, R., & K. (2018). Mine tailing disposal sites: Contamination problems, remedial options and phytocaps for sustainable remediation. *Reviews in Environmental Science and Biotechnology, 17*, 205–228. https://doi.org/10.1007/s11157-017-9453-y

Khan, W. U., Yasin, N. A., Ahmad, S. R., Ali, A., Ahmad, A., Akram, W., & Faisal, M. (2018). Role of Burkholderia cepacia CS8 in Cd-stress alleviation and phytoremediation by Catharanthus roseus. *International Journal of Phytoremediation, 20*(6), 581–592. https://doi.org/10.1080/15226514.2017.1405378

Kuffner, M., Puschenreiter, M., & Wieshammer, G. (2008). Rhizosphere bacteria affect growth and metal uptake of

heavy metal accumulating willows. *Plant and Soil, 304*, 35–44. https://doi.org/10.1007/s11104-007-9517-9

Li, F., Guo, Y., Wang, Z., & Mu, Y. (2022). Influence of different phytoremediation on soil microbial diversity and community composition in saline-alkaline land. *International Journal of Phytoremediation, 24*(5), 507–517. https://doi.org/10.1080/15226514.2021.1955240

Lin, H., Liu, C., Li, B., & Dong, Y. (2021). *Trifolium repens* L. regulated phytoremediation of heavy metal contaminated soil by promoting soil enzyme activities and beneficial rhizosphere associated microorganisms. *Journal of Hazardous Materials, 402*, 12389. https://doi.org/10.1016/j.jhazmat.2020.123829

Liu, D., An, Z., Mao, Z., Ma, L., & Lu, Z. (2015). Enhanced heavy metal tolerance and accumulation by transgenic sugar beets expressing Streptococcus thermophilus StGCS-GS in the presence of Cd, Zn and Cu alone or in combination. *PLoS One, 10*(6), e0128824. https://doi.org/10.1371/journal.pone.0128824

Liu, J., Chen, X., Shu, H. Y., Lin, X. R., Zhou, Q. X., Bramryd, T., & Huang, L. N. (2018). Microbial community structure and function in sediments from e-waste contaminated rivers at Guiyu area of China. *Environmental Pollution, 235*, 171–179. https://doi.org/10.1016/j.envpol.2017.12.008

Ma, Q., Qu, Y. Y., Zhang, X. W., Shen, W. L., Liu, Z. Y., Wang, J. W., & Zhou, J. T. (2015). Identification of the microbial community composition and structure of coal-mine wastewater treatment plants. *Microbiological Research, 175*, 1–5. https://doi.org/10.1016/j.micres.2014.12.013

Mahmood-ul-Hassan, M., Yousra, M., Saman, L., & Ahmad, R. (2020). Floriculture: Alternate non-edible plants for phytoremediation of heavy metal contaminated soils. *International Journal of Phytoremediation, 22*(7), 725–732. https://doi.org/10.1080/15226514.2019.1707772

Manoj, S. R., Karthik, C., Kadirvelu, K., Arulselvi, P. I., Shanmugasundaram, T., Bruno, B., & Rajkumar, M. (2020). Understanding the molecular mechanisms for the enhanced phytoremediation of heavy metals through plant growth promoting rhizobacteria: A review. *Journal of Environmental Management, 254*, 109779. https://doi.org/10.1016/j.jenvman.2019.109779

Meers, E., Hopgood, M., Lesage, E., Vervaeke, P., Tack, F. M. G., & Verloo, M. G. (2004). Enhanced phytoextraction: In search of EDTA alternatives. *International Journal of Phytoremediation, 6*(2), 95–109. https://doi.org/10.1080/16226510490454777

Miao, Q., & Yan, J. (2013). Comparison of three ornamental plants for phytoextraction potential of chromium removal from tannery sludge. *Journal of Material Cycles and Waste Management, 15*, 98–105. https://doi.org/10.1007/s10163-012-0095-4

Naito, K., Ebato, T., Endo, Y., & Shimizu, S. (1980). Effect of benzyladenine on δ-aminolevulinic acid synthetic ability and δ-aminolevulinic acid dehydratase: Differential responses to benzyladenine according to leaf age. *Zeitschrift für Pflanzenphysiologie, 96*, 95–102. https://doi.org/10.1016/S0044-328X(80)80047-X

Narendrula-Kotha, R., & Nkongolo, K. K. (2017). Bacterial and fungal community structure and diversity in a mining region under long-term metal exposure revealed by metagenomics sequencing. *Ecological Genetics and Genomics, 2*, 13–24. https://doi.org/10.1016/j.egg.2016.11.001

Naumof, D. G., & Dedysh, S. N. (2012). Lateral gene transfer between the Bacteroidetes and Acidobacteria: Te case of α-l-rhamnosidases. *FEBS Letters, 586*, 3843–3851.

Ni, C., Horton, D. J., Rui, J., Henson, M. W., Jiang, Y., Huang, X., & Learman, D. R. (2016). High concentrations of bioavailable heavy metals impact freshwater sediment microbial communities. *Annals of Microbiology, 66*(3), 1003–1012. https://doi.org/10.1007/s13213-015-1189-8

Novo, L. A. B., Castro, P. M. L., Alvarenga, P., & da Silva, E. F. (2018). Plant growth-promoting rhizobacteria-assisted phytoremediation of mine soils. *Bio-geotechnologies for Mine Site Rehabilitation, 1*, 281–295. https://doi.org/10.1016/B978-0-12-812986-9.00016-6. Elsevier.

Nurzhanova, A., Mukasheva, T., Berzhanova, R., Kalugin, S., Omirbekova, A., & Mikolasch, A. (2021). Optimization of microbial assisted phytoremediation of soils contaminated with pesticides. *International Journal of Phytoremediation, 23*(5), 482–491. https://doi.org/10.1080/15226514.2020.1825330

Panda, A., Patra, D. K., Acharya, S., Pradhan, C., & Patra, H. K. (2020). Assessment of the phytoremediation potential of Zinnia elegans L. plant species for hexavalent chromium through pot experiment. *Environmental Technology & Innovation, 20*, 101042. https://doi.org/10.1016/j.eti.2020.101042

Patra, D. K., Pradhan, C., & Patra, H. K. (2020). Assessment of chromium phytotoxicity, phytoremediation and tolerance potential of sesbania sesban and brachiaria mutica grown on chromite mine overburden dumps and garden soil. *Chemosphere, 252*, 126553.

Pires, C., Franco, A. R., Pereira, S. I., Henriques, I., Correia, A., Magan, N., & Castro, P. M. (2017). Metal (loid)-contaminated soils as a source of culturable heterotrophic aerobic bacteria for remediation applications. *Geomicrobiology Journal, 34*(9), 760–768. https://doi.org/10.1080/01490451.2016.1261968

Prasad, M. N. V., Greger, M., & Landberg, T. (2001). Acacia nilotica L. bark removes toxic elements from solution: corroboration from toxicity bioassay using Salix viminalis L. in hydroponic system. *International Journal of Phytoremediation, 3*(3), 289–300. https://doi.org/10.1080/15226510108500060

Qayyum, S., Khan, I., Meng, K., Zhao, Y., & Peng, C. (2020). A review on remediation technologies for heavy metals contaminated soil. *Central Asian Journal of Environmental Science and Technology Innovation, 1*(1), 21–29.

Raj, D., Kumar, A., & Maiti, S. K. (2020). Brassica juncea (L.) Czern (Indian mustard): a putative plant species to facilitate the phytoremediation of mercury contaminated soils. *International Journal of Phytoremediation, 22*(7), 733–744. https://doi.org/10.1080/15226514.2019.1708861

Rajendran, S., Priya, T. A. K., Khoo, K. S., Hoang, T. K., Ng, H. S., Munawaroh, H. S. H., & Show, P. L. (2022). A critical review on various remediation approaches for heavy metal contaminants removal from contaminated soils. *Chemosphere, 287*, 132369.

Ramana, S., Biswas, A. K., Singh, A. B., & Ahirwar, N. K. (2012). Phytoremediation of chromium by tuberose.

National Academy Science Letters, 35(2), 71–73. https://doi.org/10.1007/s40009-012-0016-z

Ramana, S., Biswas, A. K., Singh, A. B., Ahirwar, N. K., & Subba Rao, A. (2013). Potential of rose for phytostabilization of chromium contaminated soils. *Indian Journal of Plant Physiology, 18*(4), 381–383. https://doi.org/10.1007/s40502-013-0055-6

Ramana, S., Biswas, A., Singh, A., Ajay, Ahirwar, N., & Rao, A. (2015a). Tolerance of ornamental succulent plant crown of thorns (*Euphorbia milli)* to chromium and its remediation. *International Journal of Phytoremediation, 17*, 363–368. https://doi.org/10.1080/15226514.2013.862203

Ramana, S., Biswas, A. K., Singh, A. B., Ajay, Ahirwar, N. K., Prasad, R. D., & Srivastava, S. (2015b). Potential of Mauritius hemp (Furcraea gigantea Vent.) for the remediation of chromium contaminated soils. *International Journal of Phytoremediation, 17*(7), 709–715. https://doi.org/10.1080/15226514.2014.964842

Rawat, S. R., Männistö, M. K., Bromberg, Y., & Häggblom, M. M. (2012). Comparative genomic and physiological analysis provides insights into the role of Acidobacteria in organic carbon utilization in Arctic tundra soils. *FEMS Microbiology Ecology, 82*, 341–355.

Rezapour, S., Atashpaz, B., Moghaddam, S. S., Kalavrouziotis, I. K., & Damalas, C. A. (2019). Cadmium accumulation, translocation factor, and health risk potential in a wastewater-irrigated soil-wheat (*Triticum aestivum* L.) system. *Chemosphere, 231*, 579–587. https://doi.org/10.1016/j.chemosphere.2019.05.095

Sehrawat, G., Singh, R., & Kaushik, A. (2021). Tolerance of three ornamental plant species to chromium contamination in soil and their potential for phytoextraction and phytostabilization of the toxic metal. *Current World Environment, 16*(2), 386. https://doi.org/10.12944/CWE.16.2.06

Shanker, A. K., Djanaguiraman, M., Sudhagar, R., Chandrashekar, C. N., & Pathmanabhan, G. (2004). Differential antioxidative response of ascorbate glutathione pathway enzymes and metabolites to chromium speciation stress in green gram (Vigna radiata (L) R Wilczek, cv CO 4) roots. *Plant Science, 166*, 1035–1043. https://doi.org/10.1016/J.PLANTSCI.2003.12.015

Sharma, A., Kapoor, D., Wang, J., Shahzad, B., Kumar, V., Bali, A. S., & Yan, D. (2020). Chromium bioaccumulation and its impacts on plants: An overview. *Plants, 9*, 100. https://doi.org/10.3390/plants9010100

Simonsen, A., Badawi, N., Anskjær, G. G., Albers, C. N., Sørensen, S. R., Sørensen, J., & Aamand, J. (2012). Intermediate accumulation of metabolites results in a bottleneck for mineralisation of the herbicide metabolite 2, 6-dichlorobenzamide (BAM) by Aminobacter spp. *Applied Microbiology and Biotechnology, 94*(1), 237–245. https://doi.org/10.1007/s00253-011-3591-x

Song, P., Xu, D., Yue, J., Ma, Y., Dong, S., & Feng, J. (2022). Recent advances in soil remediation technology for heavy metal contaminated sites: A critical review. *Science of the Total Environment, 838*, 156417. https://doi.org/10.1016/j.scitotenv.2022.156417

Subhashini, V., & Swamy, A. V. V. S. (2013). Phytoremediation of cadmium and chromium from contaminated soils using Physalis minima Linn. *American International Journal of Research in Formal Applied and Natural Science., 3*(1), 13–260.

Suman, J., Uhlik, O., Viktorova, J., & Macek, T. (2018). Phytoextraction of heavy metals: A promising tool for clean-up of polluted environment? *Frontiers in Plant Science, 9*, 1476. https://doi.org/10.3389/fpls.2018.01476

Taneja, S., Chawla, H., Pipil, H., Yadav, S., Karaca, O., & Haritash, A. (2022). Sustainable treatment of metal-contaminated soil by electrokinetic remediation. In V. Kanwar, S. Sharma, & C. Prakasam (Eds.), *Proceedings of International Conference on Innovative Technologies for Clean and Sustainable Development (ICITCSD – 2021)* (pp. 75–84). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-93936-6_7

Taneja, S., Karaca, O., & Haritash, A. K. (2023). Combined effects of high voltage gradient and electrolyte conditioning on electrokinetic remediation for chromium (VI)-contaminated soils. *Rendiconti Lincei. Scienze Fisiche e Naturali, 34*, 635–646. https://doi.org/10.1007/s12210-023-01159-z

Tripathi, S., Singh, V. K., Srivastava, P., Singh, R., Devi, R. S., Kumar, A., & Bhadouria, R., et al. (2020). Phytoremediation of organic pollutants: current status and future directions. In P. Singh, A. Kumar, & A. B. Borthakur (Eds.), *Abatement of Environmental Pollutants* (pp. 81–105). Elsevier. https://doi.org/10.1016/B978-0-12-818095-2.00004-7

Turner, T. R., Ramakrishnan, K., Walshaw, J., Heavens, D., Alston, M., Swarbreck, D., & Poole, P. S. (2013). Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *The ISME Journal, 7*(12), 2248–2258. https://doi.org/10.1038/ismej.2013.119

urRehman, Z., Junaid, M. F., Ijaz, N., Khalid, U., & Ijaz, Z. (2023). Remediation methods of heavy metal contaminated soils from environmental and geotechnical standpoints. *Science of The Total Environment, 867*, 161468. https://doi.org/10.1016/j.scitotenv.2023.161468

Uroz, S., Oger, P., Lepleux, C., Collignon, C., Frey-Klett, P., & Turpault, M. P. (2011). Bacterial weathering and its contribution to nutrient cycling in temperate forest ecosystems. *Research in Microbiology, 162*(9), 820–831. https://doi.org/10.1016/j.resmic.2011.01.013

U.S. EPA. (1984). *Health Effects Assessment for Hexavalent Chromium (1984 Final)*. Washington, D.C.: U.S. Environmental Protection Agency. EPA/540/1-86/019 (NTIS PB86134301).

Vajpayee, P., Tripathi, R. D., Rai, U. N., Ali, M. B., & Singh, S. N. (2000). Chromium (VI) accumulation reduces chlorophyll biosynthesis, nitrate reductase activity and protein content in Nymphaea alba L. *Chemosphere, 41*(7), 1075–1082. https://doi.org/10.1016/S0045-6535(99)00426-9

Vasarevicius, S., Danila, V., & Janusevicius, T. (2020). Immobilisation of cadmium, copper, lead, and nickel in soil using nano zerovalent iron particles: Ageing effect on heavy metal retention. *Water, Air, & Soil Pollution, 231*(496), 1–11. https://doi.org/10.1007/s11270-020-04864-9

Vázquez, M. D., Poschenrieder, C. H., & Barcelo, J. (1987). Chromium VI induced structural and ultrastructural

changes in bush bean plants (Phaseolus vulgaris L.). *Annals of Botany, 59*(4), 427–438. https://doi.org/10.1093/oxfordjournals.aob.a087331

Wang, M., Liu, P., Xiong, W., Zhou, Q., Wangxiao, J., Zeng, Z., & Sun, Y. (2018). Fate of potential indicator antimicrobial resistance genes (ARGs) and bacterial community diversity in simulated manure-soil microcosms. *Ecotoxicology and Environmental Safety, 147*, 817–823. https://doi.org/10.1016/j.ecoenv.2017.09.055

Wang, T., Li, F., Lu, Q., Wu, G., Jiang, Z., Liu, S., & Sun, C. (2021). Diversity, novelty, antimicrobial activity, and new antibiotics of cultivable endophytic actinobacteria isolated from psammophytes collected from Taklamakan Desert. *Journal of Pharmaceutical Analysis, 11*(2), 241–250. https://doi.org/10.1016/j.jpha.2020.06.004

Ward, N. L., Challacombe, J. F., Janssen, P. H., Henrissat, B., Coutinho, P. M., Wu, M., & Kuske, C. R. (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology, 75*(7), 2046–2056. https://doi.org/10.1128/AEM.02294-08

Wei, Z., Hao, Z., Li, X., Guan, Z., Cai, Y., & Liao, X. (2019). The effects of phytoremediation on soil bacterial communities in an abandoned mine site of rare earth elements. *Science of the Total Environment, 670*, 950–960. https://doi.org/10.1016/j.scitotenv.2019.03.118

Wu, H., Zeng, G., Liang, J., Chen, J., Xu, J., Dai, J., & Wan, J. (2016). Responses of bacterial community and functional marker genes of nitrogen cycling to biochar, compost and combined amendments in soil. *Applied Microbiology and Biotechnology, 100*(19), 8583–8591. https://doi.org/10.1007/s00253-016-7614-5

Xia, S., Song, Z., Jeyakumar, P., Shaheen, S. M., Rinklebe, J., Ok, Y. S., & Wang, H. (2019). A critical review on bioremediation technologies for Cr (VI)-contaminated soils and wastewater. *Critical Reviews in Environmental Science and Technology, 49*(12), 1027–1078. https://doi.org/10.1080/10643389.2018.1564526

Xiao, M., Luo, Y., Zhang, H., Yu, Y., Yao, H., Zhu, Z., & Ge, T. (2022). Microplastics shape microbial communities affecting soil organic matter decomposition in paddy soil. *Journal of Hazardous Materials, 431*, 128589. https://doi.org/10.1016/j.jhazmat.2022.128589

Yan, X., Luo, X., & Zhao, M. (2016). Metagenomic analysis of microbial community in uranium-contaminated soil. *Applied Microbiology and Biotechnology, 100*(1), 299–310. https://doi.org/10.1007/s00253-015-7003-5

Yasin, N. A., Khan, W. U., Ahmad, S. R., Ahmad, A., Akram, W., & Ijaz, M. (2019). Role of Acinetobacter sp. cs9 in improving growth and phytoremediation potential of catharanthus longifolius under cadmium stress. *Polish Journal of Environmental Studies, 28*(1), 435–443. https://doi.org/10.15244/pjoes/80806

Zeng, P., Guo, Z., Cao, X., Xiao, X., Liu, Y., & Shi, L. (2018). Phytostabilisation potential of ornamental plants grown in soil contaminated with cadmium. *International Journal of Phytoremediation, 20*(4), 311–320. https://doi.org/10.1080/15226514.2017.1381939

Zhang, J., Shi, Q., Fan, S., Zhang, Y., Zhang, M., & Zhang, J. (2021). Distinction between Cr and other heavy–metal–resistant bacteria involved in C/N cycling in contaminated soils of copper producing sites. *Journal of Hazardous Materials, 402*, 123454. https://doi.org/10.1016/j.jhazmat.2020.123454

Zhang, S., Li, Y., Wang, P., Zhang, H., Ali, E. F., Li, R., & Zhang, Z. (2023). Lactic acid bacteria promoted soil quality and enhanced phytoextraction of Cd and Zn by mustard: A trial for bioengineering of toxic metal contaminated mining soils. *Environmental Research, 216*, 114646.

Zulkernain, N. H., Uvarajan, T., & Ng, C. C. (2023). Roles and significance of chelating agents for potentially toxic elements (PTEs) phytoremediation in soil: A review. *Journal of Environmental Management, 341*, 117926. https://doi.org/10.1016/j.jenvman.2023.117926

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Journal of Computational Electronics
https://doi.org/10.1007/s10825-024-02188-z

Plasma-assisted carbon nanotube for solar cell application

Suraj Kumar Singh[1] · **Ishu Sharma**[1] · **Suresh C. Sharma**[1]

Received: 3 May 2024 / Accepted: 31 May 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
This work investigated a method for improving the efficiency of solar cells through the incorporation of carbon nanotubes (CNTs), which were used as the absorber layer of the solar cell. The CNTs were generated using plasma-enhanced chemical vapor deposition (PECVD). The use of the PECVD-generated CNTs in the absorber layer of the solar cell was found to increase the electrical conductivity due to the introduction of a large number of free charge carriers in the form of electrons and holes. We were thus able for the first time to estimate a relation between plasma variables and the efficiency of the proposed solar cell. The results showed that an increase in electron and ion density resulted in an increase in the efficiency of the solar cell, whereas an increase in electron and ion temperature led to a decrease in efficiency. We also studied the variation in efficiency in relation to the absorber layer of the proposed solar cell structure. The results obtained were consistent with those from previous studies based on solar cells.

Keywords Plasma-assisted carbon nanotubes · Solar cell efficiency · Fill factor · Quantum efficiency · Efficiency of the solar cell equipped with CNTs

Abbreviations

A	Effective area of the solar cell
R_s	Series resistance
S_e	Surface recombination velocity for electrons
N	Electron concentration
J	Current density
R_p	Recombination rate of holes
n_o	Ideality factor
R_{sh}	Shunt resistance in parallel
J_o	Reverse saturation current density
CNTs	Carbon nanotubes
R_n	Recombination rate of electrons
L	Length of CNT
R	Radius of CNT
PCE	Power conversion efficiency
T_{eo}	Electron temperature
T_{io}	Ion temperature
λ_D	Debye length
P_{in}	Solar irradiance
G_n	Generation rate of electrons
T	Absolute temperature
q	Electronic charge
n_{eo}	Electron density
S_h	Surface recombination velocity for holes
n_{io}	Ion density
N	No. of carbon nanotubes (CNTs)
η	Efficiency of solar cell
P_{out}	Power output
ITO	Indium tin oxide
V_{OC}	Open-circuit voltage
J_n	Electron current density
G_p	Generation rate of holes
AM	Air mass spectrum
J_{SC}	Photocurrent density
CB	Conduction band
K_B	Boltzmann constant
VB	Valence band
p	Hole concentration
φ	Metal work function
TCO	Transparent conductive oxide
J_p	Hole current density
QE	Quantum efficiency
V_S	Surface potential

✉ Suresh C. Sharma
suresh321sharma@gmail.com

Suraj Kumar Singh
singhsurajkumar723@gmail.com

Ishu Sharma
ishusharma202122@gmail.com

[1] Department of Applied Physics, Delhi Technological University, Delhi 110042, India

Published online: 01 July 2024

t Thickness of absorber layer
FF Fill factor

1 Introduction

A solar cell is a photovoltaic device that converts light radiation into electrical energy [1]. It works the same way as a p–n-based photodetector. When sunlight is incident on the solar cell, the electrons in the conduction band (CB) jump to the valence band (VB), thus creating holes in the CB that facilitate the flow of current in the solar cell [2]. This electron- and hole-generated current is called photocurrent [3–7]. The photovoltaic conversion efficiency (PCE) of the solar cell has been reported up to 1–5% [1, 8], depending on the photocurrent density, fill factor (FF), input power irradiance (P_{in}), and open-circuit voltage (V_{OC}) [9]. There are various types of solar cells, including thin-film, perovskite, and dye-sensitized solar cells [2]. Perovskite-based solar cells have reported PCE up to 22.1% [10], but they are not stable under light illumination conditions. Researchers have reported that when carbon nanotubes (CNTs) are used in solar cells, an increase of 1.3% to as much as 30% in PCE has been achieved [9]. The structure of CNT-based solar cells is the same as the single p–n junction solar cell, where the CNTs can behave as both p- and n-type semiconductors [2, 10]. CNTs were first grown by the Japanese scientist Sumio Ijima in 1991 [11–13]. They are highly conductive—even higher than copper—with an extremely large surface area and high strength because of the covalent sp^2 bond between individual carbon atoms [14]. The bandgap of CNTs varies from 0.1 eV to 2 eV [14–17]. Their basic structure can be categorized into three types, i.e., armchair, zigzag, and chiral CNTs, with stability in the order armchair > zigzag > chiral [18–21]. Several kinds of solar cells with good photovoltaic properties have been developed using carbon nanotubes. We have used the plasma-enhanced chemical vapor deposition (PECVD) process for growing CNTs [22–26]. Although numerous works have investigated the use of CNTs in solar cells, no study has proposed a mathematical expression for calculating the efficiency of the solar cells based on plasma variables/parameters. In the present work, PECVD-grown CNTs were used in the absorber layer of the solar cell via simulation and a mathematical formula was devised for computing the solar cell efficiency as a function of plasma parameters including electron and ion temperature and density. With the use of CNTs, we observed an increase in the efficiency of ITO-/CdS-/CNT-based solar cells of up to 23–24%, which was similar to simulated and past research results [10]. For the simulation, we used SCAPS-1D software to construct the solar cell and calculate its efficiency. In the simulation setup, CNTs were used in the absorber layer, and the thickness of the absorber layer was varied from 100 to 1000 nm. Related variables including FF, V_{OC}, short-circuit current density (J_{SC}), PCE, and bandgap were determined. After simulation, we calculated the efficiency of the solar cell from our estimated expression, and the results were very close to the efficiency from the simulation, as shown in Fig. 11, and were also consistent with results from many previous studies [25–29, 54–70].

2 Simulation work and mathematical modeling proposed for ITO-, CdS-, and CNT-based solar cell structure

2.1 Work (1)

2.1.1 Simulation of ITO-, CdS-, and CNT-based solar cells

2.1.1.1 Solar cell simulation and incorporation of CNTs in solar cells Here, we used SCAPS-1D 3.3.10 software for the ITO-/CdS-/CNT-based solar cell simulations. In Fig. 1, CdS is illustrated in a schematic of the solar cell, acting as a buffer layer between the CNT absorber layer and the ITO window layer, with dimensions of 0.441 cm × 0.440 cm × (202 to 1102) nm, as given in Figs. 1 and 2. These fundamental components allow one to determine the electrical and photoelectric characteristics of a photovoltaic solar cell. In the proposed solar cell, the ITO and absorber layers are connected with Bi and Au material as electrical contacts [25–29].

SCAPS-1D software, which was developed by the University of Ghent, Belgium [30], can analyze the parameters of solar cells including current density, bandgap, layer thickness, quantum efficiency, fill factor, and dark saturation current. Figure 2 shows a schematic of the solar cell structure generated from SCAPS-1D, in which sunlight (incident

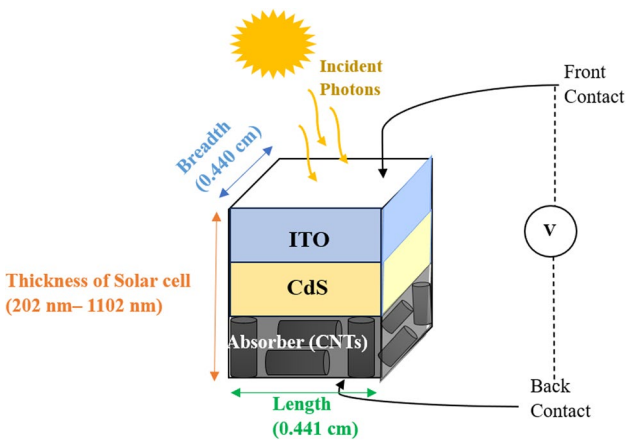


Fig. 1 Schematic diagram of ITO-/CdS-/CNT-based solar cell

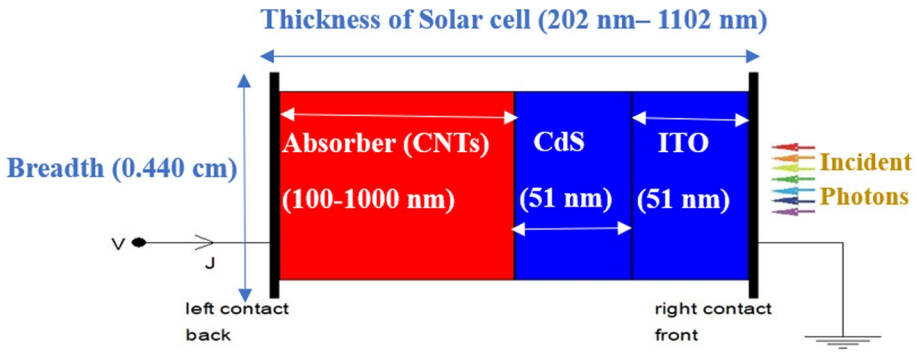


Fig. 2 Schematic diagram of solar cell constructed with SCAPS-1D software

photons) falls from the side of the right contact (front) (Bi) to the left contact (back) (Au).

SCAPS-1D uses basic concepts of physics and mathematics for semiconductor devices, as expressed in equations (a–c), as follows [25, 31]:

Poisson's equation:

$$\mathrm{div}(\varepsilon.\nabla\Psi) = -\rho \tag{1a}$$

The continuity equation for electron carriers:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\mathrm{div}(J_n) + G_n - R_n \tag{2b}$$

The continuity equation for hole carriers:

$$\frac{\partial p}{\partial t} = \frac{1}{q}\mathrm{div}(J_p) + G_p - R_p \tag{2c}$$

In the SCAPS-1D simulation software, we have assumed air mass spectrum AM 1.5 radiation, working temperature of 300 K, and an ideal solar irradiance of 1 kW m^{-2}, using an applied voltage of 0.77 V. The values of the shunt and series resistance are 5.0×10^{10} Ω cm^2 and 0.44 Ω cm^2, respectively [9, 10]. The frequency of incident light is 1×10^6 Hz.

2.1.2 Simulation results and discussion

Tables 1, 2, and 3 present the data used in the SCAPS-1D simulations, with the results shown in Figs. 3, 4, 5, and 6.

2.1.2.1 Bandgap explanation Figure 3 presents an energy band diagram illustrating the energy levels of various layers of a solar cell derived from SCAPS-1D for analysis of the device performance. The absorber layer is made of CNTs, and we know that when we increase the radius of the CNTs, the energy bandgap will decrease, as clearly shown in Fig. 3 [52]. To understand this relation between the bandgap and radius of CNTs, we can refer to a basic concept of quantum mechanics called quantum confinement. Quantum confinement processes cause the bandgap of nanoparticles to increase with decreasing particle size. The continuous energy bands control the band structure and electronic characteristics of larger bulk materials. However, the bandgap increases when a material decreases to the nanoscale because separate energy levels are formed by the confinement of electrons and holes in a very small volume. This phenomenon is understood by the concept of quantum confinement, which results from the confinement of charge

Table 1 Simulation values for the proposed solar cell layers for SCAPS-1D software [25–29]

Parameters	ITO	CdS	Absorber
Layer thickness (μm)	0.051	0.051	0.1–1
Shallow uniform acceptor density NA (cm^{-3})	–	–	1.1×10^{17}
Electron mobility (cm^2 V s^{-1})	51	170	8.1×10^4
Layer electron affinities (eV)	4.5	4.2	4.28
Layer relative dielectric permittivity	8.8	8.72	3.5
Electron thermal velocity in the layer (cm s^{-1})	1.1×10^7	1.1×10^7	1.1×10^7
Layer density of defects (cm^{-3})	–	1.0×10^{17}	1.0×10^{14}
Layer bandgap (eV)	3.5	2.3	1.2
Hole thermal velocity in the layer (cm s^{-1})	1.1×10^7	1.1×10^7	1.1×10^7
VB layer, effective density of states (cm^{-3})	1.9×10^{19}	1.9×10^{19}	6.1×10^{17}
Hole movement within the layer (cm^2 V s^{-1})	11	16	2.1×10^3
Shallow uniform donor density ND (cm^{-3})	1.1×10^{21}	1.1×10^{17}	–
CB layer, effective density of states (cm^{-3})	2.3×10^{18}	2.3×10^{18}	5.1×10^{16}

Table 2 Proposed solar cell from SCAPS-1D software simulation front and back metal contact parameters [25, 32–34]

Parameter	Right contact (front)	Left contact (back)
Electron surface recombination velocity (cm s^{-1})	1.0×10^{7}	1.0×10^{7}
Work function of used metal (eV)	4.34	5.42
Hole surface recombination velocity (cm s^{-1})	1.0×10^{7}	1.0×10^{7}

Table 3 Defects in solar cell layers [25]

Parameter	Absorber	ITO
Holes, capture cross-section (cm^2)	10^{-15}	10^{-15}
Reference for defect energy level Et	Above EV (SCAPS < 2.7)	Above EV (SCAPS < 2.7)
Energy distribution	Single type	Single type
Defect nature	Neutral	Neutral
Electrons, capture cross-section (cm^2)	10^{-15}	10^{-15}
Energy level relative to the reference (eV)	0.600	0.600
N_t total (cm^{-3})	10^{14}	10^{14}

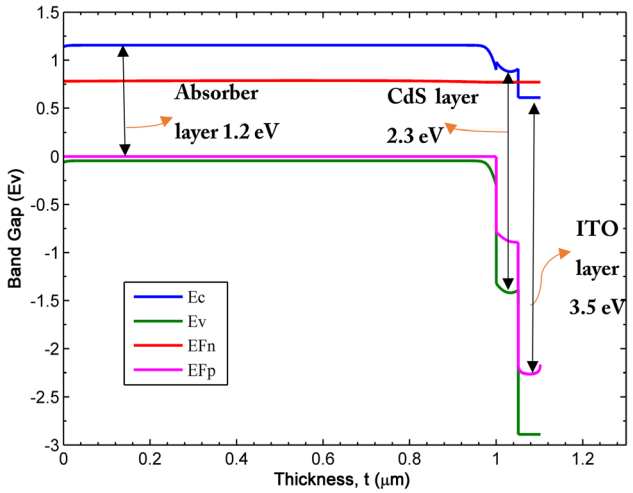


Fig. 3 Energy bandgap in relation to thickness of the absorber layer (t) of the proposed solar cell

carriers in a volume equal to their wavelength. Energy levels can be quantified when the spatial confinement of electrons and holes increases with decreasing nanoparticle size. The energy difference between the valence and conduction bands increases as a result of this quantization of energy levels.

2.1.2.2 Examination of QE curve, generated current density, and carrier generation and recombination When sunlight falls on a semiconductor, electron and hole generation and recombination take place inside the semiconductor [35–38, 38–51]. Since the absorber layer consists of CNTs which are electrically similar to semiconductors, the same phenomena of generation and recombination take place from this incident light process on the solar cell. The recombination and generation with respect to the absorption layer thickness (t) can be seen in Fig. 4a. This process facilitates the movement of charge carrier holes and electrons, which in turn causes the generation of current [35–38, 38–51], as can be seen in Fig. 4b. Similarly, we can see the quantum efficiency of the device with respect to various wavelengths of light in Fig. 4c; QE decreases with increasing wavelength because longer-wavelength photons have less energy, resulting in lower levels of carrier generation and recombination.

2.1.2.3 *J–V* characteristics of the solar cell with varied absorption layer thickness We know that for the solar cell, the power generated by the cell is $-(V\times I)$, and the power consumed by the cell is $+(V\times I)$. The photovoltaic process begins when the cell is exposed to the sun at AM 1.5. The absorption of light creates charge carriers, i.e., free electrons and holes in the absorption layer, which produces a continuous photocurrent [25]. Since CNTs have very high current flow and very low resistance, they also have many better electrical properties such as conductivity and mobility, which increase the current generation, thus enhancing the efficiency of the device. As shown in Fig. 5, increased absorption layer thickness enhances the photocurrent generated by the constructed solar cell, which indicates that as the power output of the cell increases, its efficiency will also increase. Here, we have taken the thickness of the absorber layer in the range from 100 to 1000 nm, which yields efficiency of 12–27%, consistent with the results of previous research [10, 33, 34].

2.1.2.4 SCAPS-1D-simulated solar cell with varied absorber layer thickness As shown in Fig. 6a, b, and d, in the simulated ITO/CdS/CNT solar cell, because the absorber layer is filled with a large number of CNTs, the rates of recombination and generation increase when the absorber layer thickness (t) increase, and the resulting increase in current generation leads to enhanced efficiency

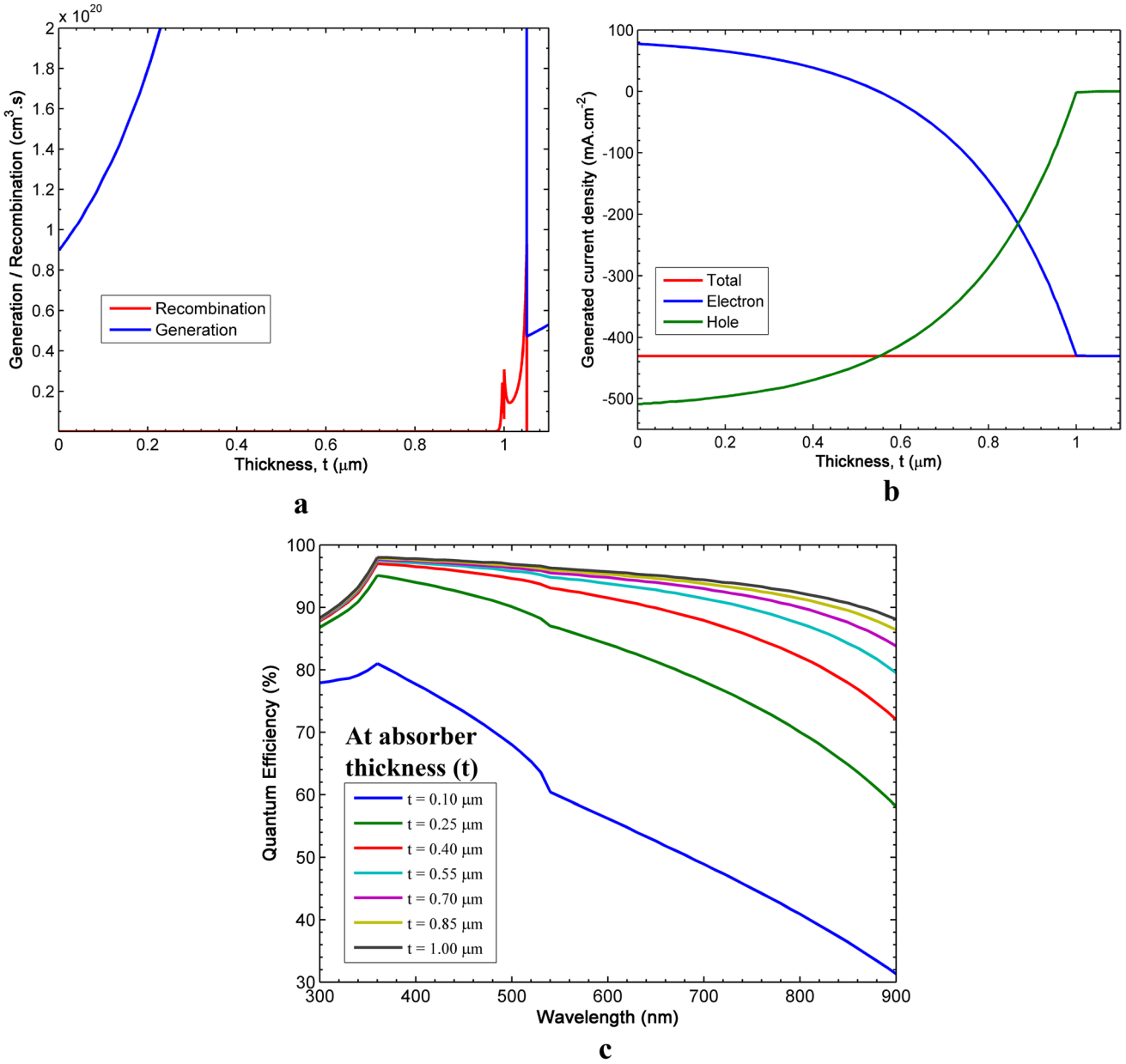


Fig. 4 Generation/recombination of carriers, generated current density, and QE of the SCAPS-1D-simulated solar cell. **a** Recombination and generation of carriers in relation to absorber layer thickness (*t*). **b** Generated current density with relation to *t*. **c** QE of the proposed simulated solar cell for incident photons with various wavelengths at varying values of *t*

of the constructed solar cell. In Fig. 6c, the parameter FF is found to decrease due to the increased generation and recombination of carriers, resulting in larger values of J_{SC} and V_{OC}. Because J_{SC} and V_{OC} are inversely proportional to FF [10], the FF value decreases. If we explain this phenomenon from a theoretical point of view, we know that a solar cell is a p–n junction device which works on a basic physics principle, namely the photovoltaic effect, where carriers are generated, which leads to the generation of electric current, and the special optical characteristics of CNTs allow the solar cell to absorb a wide range of light, including visible and near-infrared wavelengths. Increased absorption of sunlight by CNTs in the absorber layer results in increased creation of electron–hole pairs, also called excitons, since CNTs have high charge carrier mobility. In addition, the current loss decreases, which results in high J_{SC} and V_{OC} values and thus a decrease in the FF of the solar cell.

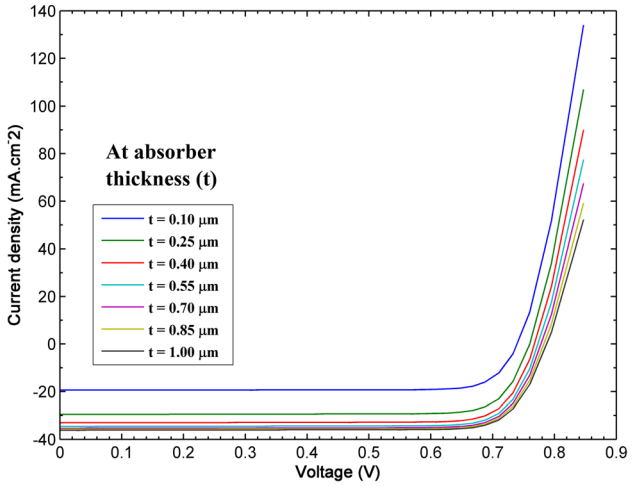


Fig. 5 Current–voltage (*J*–*V*) characteristic of the simulated solar cell with absorber layer thickness (*t*) ranging from 0.1 to 1.0 µm

2.1.3 Summary of results from SCAPS-1D simulation

The results obtained from the simulation are presented in Table 4.

2.2 Work (2)

2.2.1 Analytical modeling of ITO-, CdS-, and CNT-based solar cells

The efficiency of a solar cell can be expressed as follows [9, 10]:

$$\text{Efficiency of solar cell} = \frac{\text{Power output}\left(P_{\text{out}}\right)}{\text{Power input}\left(P_{\text{in}}\right)}$$

Solar cell efficiency (η) can be expressed mathematically as [9, 10]

$$\eta = \frac{\text{FF} \times V_{\text{OC}} \times J_{\text{SC}}}{P_{\text{in}}} \times 100 \tag{1}$$

The photocurrent density of the solar cell is defined as [10]

$$J_{\text{SC}} = J_{\text{O}}\left\{\exp\left(\frac{q\left(V_{\text{OC}} - \text{JAR}_{\text{S}}\right)}{n_{\text{o}}K_{\text{B}}T}\right) - 1\right\} + \left(\frac{V_{\text{OC}} - \text{JAR}_{\text{S}}}{\text{AR}_{\text{SH}}}\right) - J \tag{2}$$

Now, substituting the J_{SC} from Eq. (2) into Eq. (1), we obtain

$$\eta = \left(\frac{\text{FF} \times V_{\text{OC}}}{P_{\text{in}}}\right)\left[J_{\text{O}}\left\{\exp\left(\frac{q\left(V_{\text{OC}} - \text{JAR}_{\text{S}}\right)}{n_{\text{o}}K_{\text{B}}T}\right) - 1\right\} + \left(\frac{V_{\text{OC}} - \text{JAR}_{\text{S}}}{\text{AR}_{\text{SH}}}\right) - J\right] \times 100 \tag{3}$$

and power output will be

$$P_{\text{out}} = \left(\text{FF} \times V_{\text{OC}}\right)\left[J_{\text{O}}\left\{\exp\left(\frac{q\left(V_{\text{OC}} - \text{JAR}_{\text{S}}\right)}{n_{\text{o}}K_{\text{B}}T}\right) - 1\right\} + \left(\frac{V_{\text{OC}} - \text{JAR}_{\text{S}}}{\text{AR}_{\text{SH}}}\right) - J\right] \tag{4}$$

Here, we assume that the front face of the solar cell is a rectangular structure, as given in Fig. 1. Thus, the area of the proposed solar cell = side length × side width.

Therefore, the area of the solar cell can be expressed as

$$A = N \times (2R) \times L \tag{5}$$

where N is the available number of CNTs within the absorber layer, and the CNT radius and length are denoted by R and L, respectively.

Since we have used CNTs in the solar cell as the absorber layer, and the volume of this layer will vary with respect to CNT radius (R) and length (L) and the number of CNTs (N), the only quantity that will change is the number of CNTs when we alter that the value of t, maintaining the CNT length and radius constant. As indicated in Fig. 7, the absorber length and width are 0.440 and 0.441 cm, respectively, and its thickness (t) varies from 100 to 1000 nm.

For calculating the number of CNTs (N), we use the equal-volume concept, i.e., the volume of the absorber layer is equal to the available number of CNTs multiplied by the volume of a single CNT, as follows:

$$\text{Volume of absorber layer} = N \times \text{volume of single CNT}$$

Since CNTs are cylindrical, the volume of a single CNT will be $= (\pi \times \text{R}^2 \times \text{L})$

and the volume of the absorber layer will be expressed as Length × Breadth × Thickness of absorber layer

Thus,

$$N = \frac{\text{Volume of the absorber layer}}{\text{Volume of single CNT}} = \frac{\text{Length} \times \text{Breadth} \times \text{Thickness of the absorber layer}}{\pi \times R^2 \times L}$$

Now, from Eq. (5), substituting the value of A from Eqs. (3) and (4), we obtain

$$\eta = \left(\frac{\text{FF} \times V_{\text{OC}}}{P_{\text{in}}}\right)\left[J_{\text{O}}\left\{\exp\left(\frac{q\left(V_{\text{OC}} - \text{NL}(2R)\text{JR}_{\text{S}}\right)}{n_{\text{o}}K_{\text{B}}T}\right) - 1\right\} + \left(\frac{V_{\text{OC}} - \text{NL}(2R)\text{JR}_{\text{S}}}{\text{NL}(2R)R_{\text{SH}}}\right) - J\right] \times 100 \tag{6}$$

and power output will be

$$P_{\text{out}} = \left(\text{FF} \times V_{\text{OC}}\right)\left[J_{\text{O}}\left\{\exp\left(\frac{q\left(V_{\text{OC}} - \text{NL}(2R)\text{JR}_{\text{S}}\right)}{n_{\text{o}}K_{\text{B}}T}\right) - 1\right\} + \left(\frac{V_{\text{OC}} - \text{NL}(2R)\text{JR}_{\text{S}}}{\text{NL}(2R)R_{\text{SH}}}\right) - J\right] \tag{7}$$

The radius of the CNT can be expressed in terms of the plasma parameters [52, 53], or

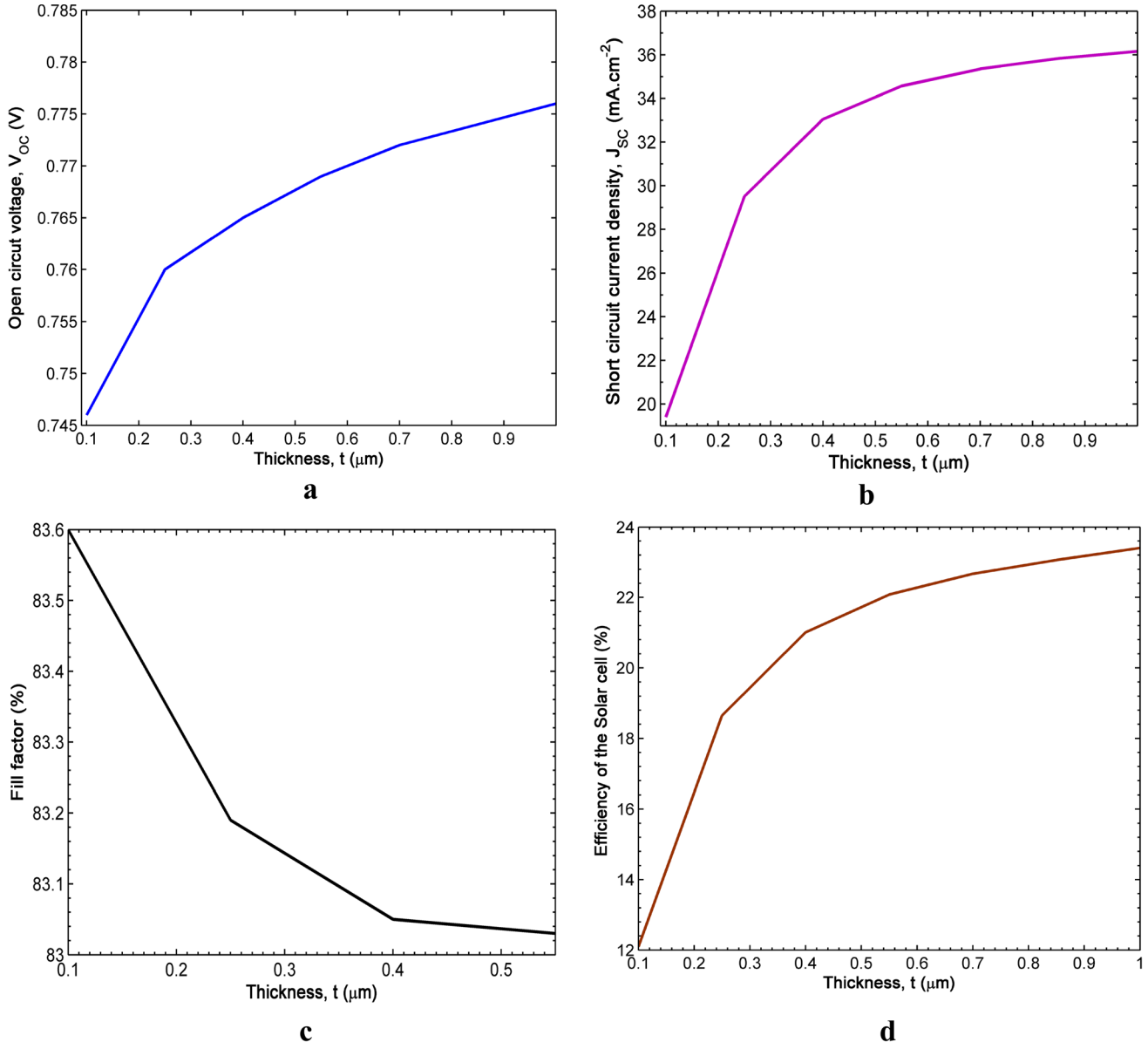


Fig. 6 Efficiency variables of the simulated structure of the solar cell in SCAPS-1D in relation to the thickness of the absorber layer (*t*). **a** V_{OC} in relation to *t*. **b** J_{SC} in relation to *t*. **c** FF in relation to *t*. **d** Efficiency in relation to *t*

Table 4 Summary of the results from SCAPS-1D simulation of the constructed solar cell

Thickness of absorber layer (*t*) (μm)	Volume of absorber layer (cm^3)	Total thickness of solar cell (μm)	FF (%)	V_{OC} (V)	J_{SC} (mA cm^{-2})	η (%)
0.10	1.94×10^{-8}	0.202	83.60	0.746	19.406	12.10
0.25	4.85×10^{-8}	0.325	83.19	0.760	29.507	18.65
0.40	7.76×10^{-8}	0.502	83.05	0.765	33.049	21.01
0.55	10.6×10^{-8}	0.652	83.03	0.769	34.571	22.08
0.70	13.5×10^{-8}	0.802	83.06	0.772	35.359	22.67
0.85	16.4×10^{-8}	0.952	83.11	0.774	35.835	23.06
1.00	19.0×10^{-8}	1.102	83.35	0.776	36.161	23.40

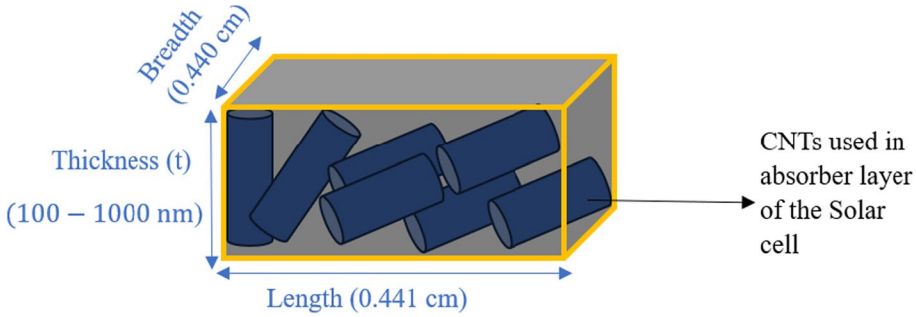


Fig. 7 CNTs utilized in solar cell absorber layer

$$R^2 = \exp\left(\frac{V_S L}{q}\right)\left[\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}} + \frac{n_{io}}{T_{io}}\right)}\right]$$

$$R = \exp\left(\frac{V_S L}{2q}\right)\left[\sqrt{\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}} + \frac{n_{io}}{T_{io}}\right)}}\right] \tag{8}$$

Now, substituting the value of R from Eq. (8) into Eqs. (6) and (7), we obtain

$$\eta = \left(\frac{\mathrm{FF}\times V_{OC}}{P_{in}}\right)\left[\begin{array}{c} J_O\left\{\exp\left(\frac{q\left(V_{OC}-2\mathrm{NL}\exp\left(\frac{V_S L}{2q}\right)\left[\sqrt{\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}}+\frac{n_{io}}{T_{io}}\right)}}\right]JR_S\right)}{n_o K_B T}\right)-1\right\}+ \\ \left(\frac{V_{OC}-2\mathrm{NL}\exp\left(\frac{V_S L}{2q}\right)\left[\sqrt{\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}}+\frac{n_{io}}{T_{io}}\right)}}\right]JR_S}{2\mathrm{NL}\exp\left(\frac{V_S L}{2q}\right)\left[\sqrt{\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}}+\frac{n_{io}}{T_{io}}\right)}}\right]R_{SH}}\right)-J \end{array}\right]\times 100 \tag{9}$$

and

$$P_{out} = (\mathrm{FF}\times V_{oc})\left[\begin{array}{c} J_O\left\{\exp\left(\frac{q\left(V_{OC}-2\mathrm{LN}\exp\left(\frac{V_S L}{2q}\right)\left[\sqrt{\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}}+\frac{n_{io}}{T_{io}}\right)}}\right]JR_S\right)}{n_o K_B T}\right)-1\right\}+ \\ \left(\frac{V_{OC}-2\mathrm{LN}\exp\left(\frac{V_S L}{2q}\right)\left[\sqrt{\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}}+\frac{n_{io}}{T_{io}}\right)}}\right]JR_S}{2\mathrm{NL}\exp\left(\frac{V_S L}{2q}\right)\left[\sqrt{\frac{1}{4\pi q^2\left(\frac{n_{eo}}{T_{eo}}+\frac{n_{io}}{T_{io}}\right)}}\right]R_{SH}}\right)-J \end{array}\right] \tag{10}$$

2.2.2 Analytical results

Our work was conducted in two parts: the first to fabricate CNTs using the PECVD method, and the second to use these CNTs in the solar cell absorber layer. For analytical calculation, we use the values of V_{OC}, J_{SC}, J_o, n_o, R_S, R_{SH}, and P_{in} at varied absorber layer thickness, which were obtained from the SCAPS-1D simulation, as summarized in Table 4. For mathematical calculation and results, we use constant values taken from the SCAPS 1-D simulation setup, which are given in Table 5.

2.2.2.1 Solar cell efficiency in relation to plasma parameters (n_{eo}, n_{io}, T_{eo}, and T_{io})

To analytically determine the effects of plasma parameters on solar cell efficiency, we take the required PECVD data from Table 5 and from the work (1) section in Table 6 [52, 53], which are essential for solving Eqs. (6), (7), and (9) to obtain the efficiency. When we put all the parameter values into Eqs. (9) and (10), we obtain the solar cell power output and efficiency, as shown in Table 6 in the work (2) section.

2.2.2.2 Influence of plasma parameters electron and ion density (n_{eo} and n_{io}) on solar cell efficiency

Figure 8 presents plots of the solar cell power output and efficiency in relation to electron and ion density (n_{eo} and n_{io}). In Fig. 8a–d, we see that when we increase the electron and ion density, the output power of the proposed solar cell increases. This is because, as we know, when we increase the plasma parameters n_{eo} and n_{io}, the radius of the CNTs will decrease, which leads to an increase in the voltage output of the device [52, 53], which in turn leads to an increase in the power output and efficiency of the solar cell. To understand this phenomenon, the increased electron density (n_{eo}) and ion density (n_{io}) parameters in the PECVD process typically result in a larger concentration of reactive species and energy in the plasma. The deposited CNTs and growth dynamics may be influenced by this higher energy and reactivity. Because higher plasma density and input power facilitate more rapid nucleation and growth of CNTs, the radius of the CNTs decreases

Table 5 Constant values for mathematical calculation

Parameter	Value
P_{in}	1000 W m^{-2}
q	1.6×10^{-19} C
R_S	0.44 Ω cm^2
R_{Sh}	5.0×10^{10} Ω cm^2
K_b	1.38×10^{-23} j K^{-1}
T	300 K
n_o	2.1

Table 6 Efficiency of solar cell calculated from Eqs. (6), (7), and (9)

Work (1)						Work (2)						
Results from PECVD method						Results from simulation and analytical work						
n_{eo} (cm^{-3})	n_{io} (cm^{-3})	T_{eo} (K)	T_{io} (K)	R (nm)	L (μm)	t (μm)	N	V_{OC} (v)	FF (%)	J ($mA.cm^{-2}$)	J_O ($mA.cm^{-2}$)	η (%)
1.0×10^{7}	5.0×10^{6}	5800	2400	2.0	0.1	0.10	1.5×10^{12}	0.746	83.60	18.59	0.45×10^{-4}	12.42
5.0×10^{6}	4.2×10^{6}	5220	2300	3.0	0.1	0.25	1.7×10^{12}	0.760	83.19	28.22	0.49×10^{-4}	18.89
1.0×10^{6}	2.5×10^{6}	4640	2200	4.0	0.1	0.40	1.5×10^{12}	0.765	83.05	31.63	0.50×10^{-4}	21.22
1.0×10^{5}	1.0×10^{6}	3480	2100	5.0	0.1	0.55	1.3×10^{12}	0.769	83.03	33.09	0.49×10^{-4}	22.67
5.0×10^{8}	5.0×10^{7}	12,760	2180	5.0	1.0	0.70	1.7×10^{11}	0.772	83.06	33.83	0.47×10^{-4}	22.88
1.0×10^{9}	5.0×10^{8}	13,920	2210	5.0	2.0	0.85	1.0×10^{11}	0.774	83.11	34.26	0.46×10^{-4}	23.37
5.0×10^{9}	1.0×10^{9}	15,080	2250	5.0	3.0	1.00	8.2×10^{10}	0.776	83.35	34.66	0.45×10^{-4}	23.91

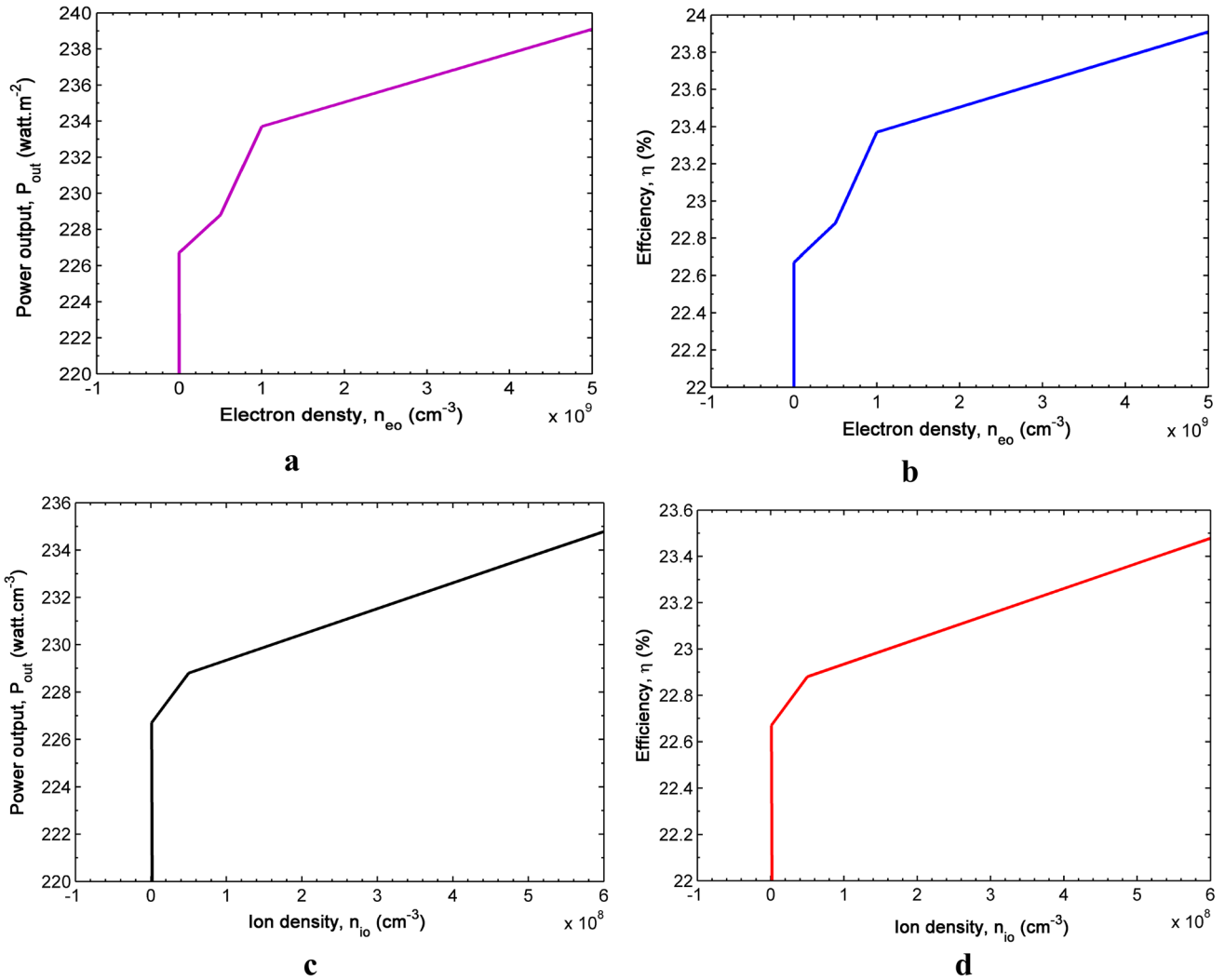


Fig. 8 Solar cell efficiency and power output in relation to plasma parameters electron and ion density (n_{eo} and n_{io}). **a** Power output of the solar cell in relation to electron density (n_{eo}). **b** Solar cell efficiency in relation to electron density (n_{eo}). **c** Power output of the solar cell in relation to ion density (n_{io}). **d** Solar cell efficiency in relation to ion density (n_{io})

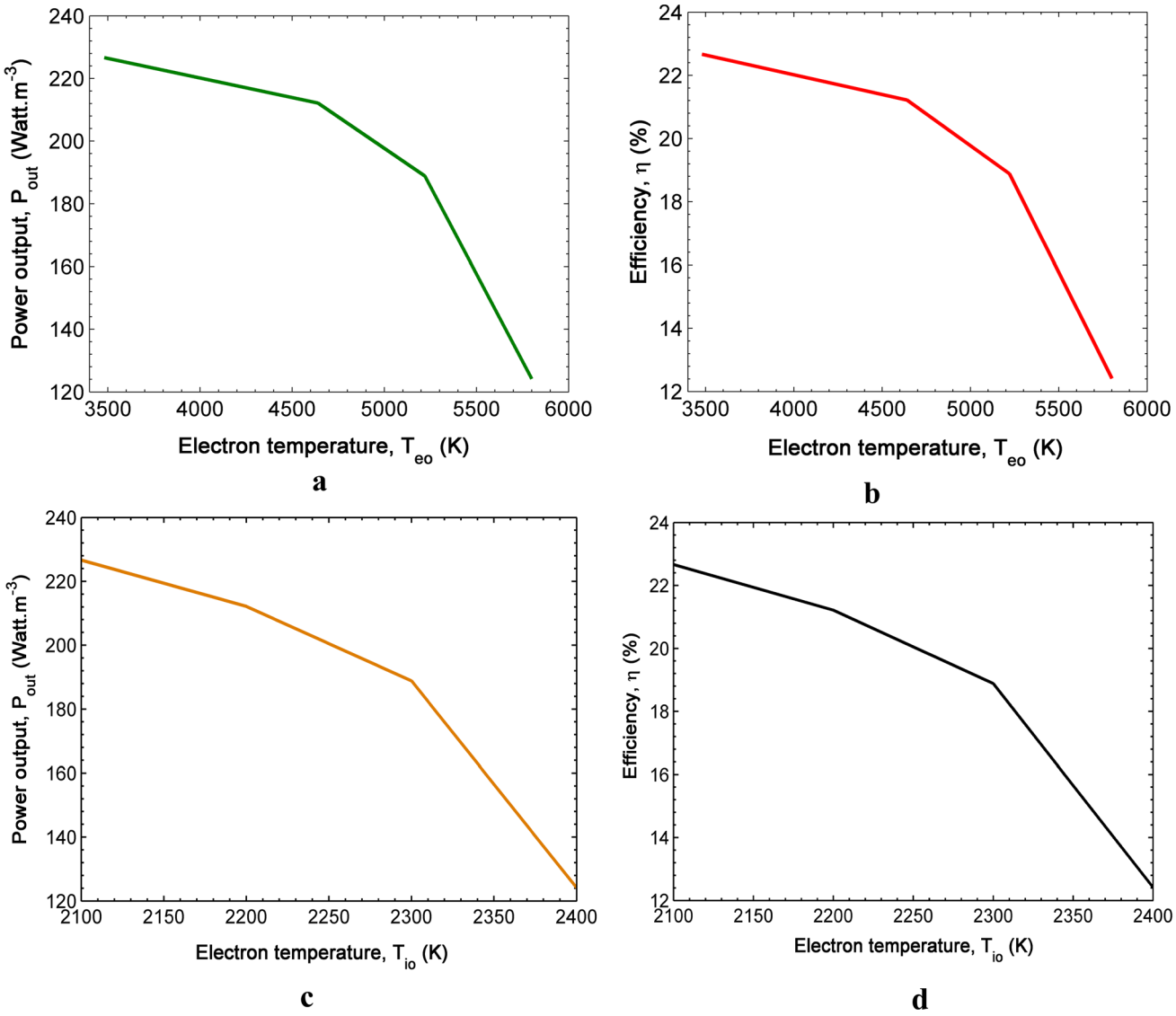


Fig. 9 Solar cell efficiency and power output in relation to plasma parameters electron and ion temperature (T_{eo} and T_{io}). **a** Solar cell output power in relation to electron temperature (T_{eo}). **b** Efficiency of the solar cell in relation to electron temperature (T_{eo}). **c** Power output of the solar cell in relation to ion temperature (T_{io}). **d** Efficiency of the solar cell in relation to ion temperature (T_{io})

with increasing plasma parameters, thus producing thinner CNTs. In addition, the etching effect, in which reactive species in the plasma may degrade the CNT sidewalls during growth, can be improved by increasing the plasma parameter values. Thinner CNTs can be obtained by this etching action, which contributes to decreasing the CNT radius. At the interface between the CNTs and the active material (e.g., buffer layer and ITO), smaller-diameter CNTs can provide a stronger electric field. This results in higher voltage output, facilitating more effective separation of photogenerated electron–hole pairs. Moreover, smaller CNTs can provide a more direct route for charge carriers to reach the electrode, hence reducing recombination losses and increasing charge accumulation efficiency. This can result in higher current, which by Ohm's law can increase the voltage output because of changes induced by the plasma treatment, so the effective surface area of the CNTs can increase while the radius decreases. Because of

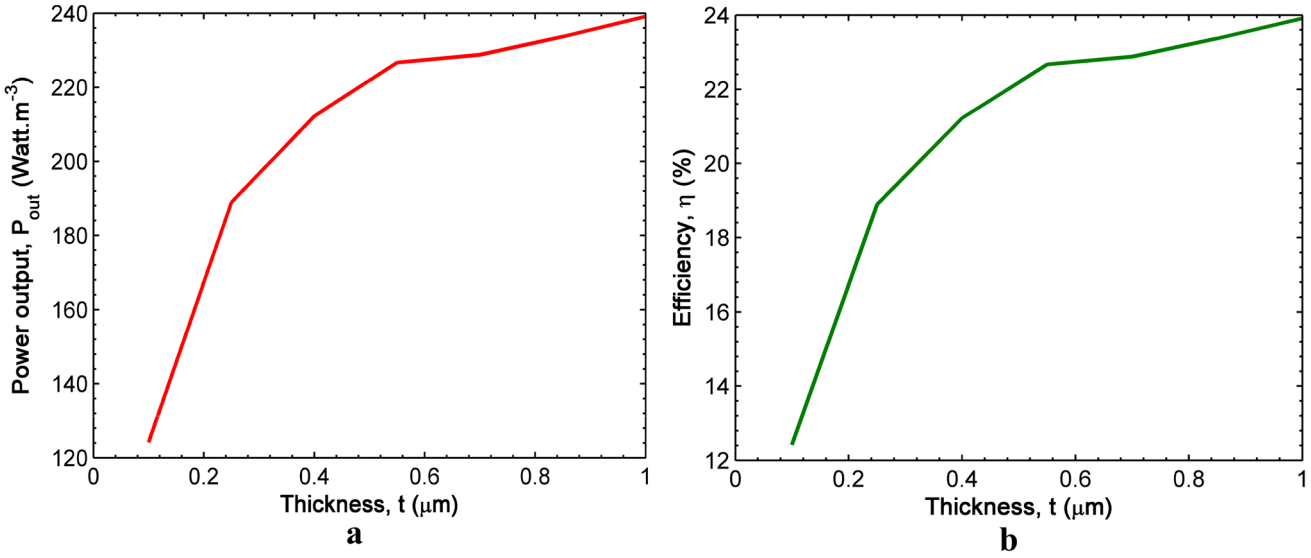


Fig. 10 Power output and efficiency of ITO-/CdS-/CNT-based solar cell influenced by absorber layer thickness (t). **a** Output power of the solar cell with respect to t. **b** Solar cell efficiency in relation to t

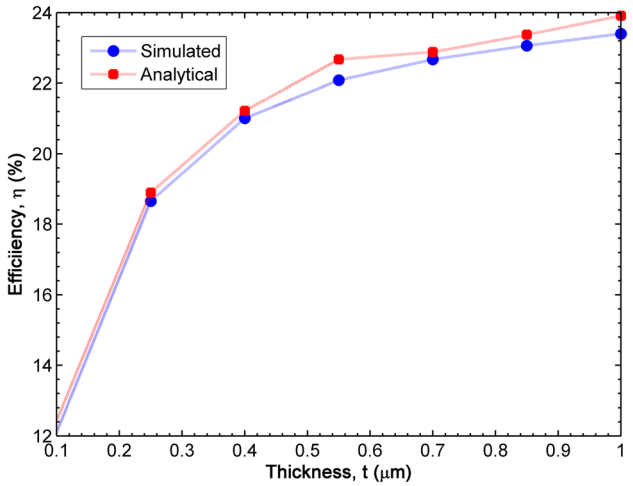


Fig. 11 Efficiency of the constructed solar cell determined from analytical and simulation work

Table 8 Summary table for impact of the different factors on solar cell efficiency

Parameter	Effect of parameters effect on solar cell efficiency
N_{eo}	Increases with increasing value of n_{eo}
N_{io}	Increases with increasing value of n_{io}
T_{eo}	Decreases with increasing value of T_{eo}
T_{io}	Decreases with increasing value of T_{io}
t	Increases with increasing value of t

this more active region for charge generation and photon absorption, we obtain higher current and voltage production.

2.2.2.3 Effect of electron and ion temperature (T_{eo} and T_{io}) on solar cell efficiency Figure 9 illustrates solar cell power output and efficiency in relation to electron and ion temperature (T_{eo} and T_{io}). As shown in Fig. 9a–d, when we

Table 7 Percentage error between simulated and analytical results

Efficiency of the Solar cell		Error (%)	Average (%)
Simulated result (%)	Analytical result (%)	$\left\lvert\frac{\text{SimulatedValue}-\text{AnalyticalValue}}{\text{SimulatedValue}}\right\rvert \times 100$	$\frac{\text{Totalsumoferror}}{\text{No.oftotalerror}}$
12.10	12.42	2.64	$\frac{2.64+1.27+0.99+2.67+0.92+1.34+2.17}{7}$
18.65	18.89	1.27	$=1.71\%$
21.01	21.22	0.99	
22.08	22.67	2.67	
22.67	22.88	0.92	
23.06	23.37	1.34	
23.40	23.91	2.17	

increase the electron and ion temperature, the output power of the solar cell decreases. This because, as we know, when T_{eo} and T_{io} are increased, the radius of the CNTs will increase, which decreases the voltage output of the device [52, 53], in turn leading to a decrease in the power output and efficiency of the solar cell. This phenomenon can be understood as follows: During synthesis procedures like PECVD, electron and ion temperature can affect CNT growth and stability. Faster growth rates and structural changes in the generated CNTs can be induced by higher temperature. However, the particular impact on the radius of CNTs can vary according to a number of variables, including the type of catalyst, the synthesis technique, the growth environment, and the kind of CNT (single-walled, multi-walled, etc.) generated. With higher electron and ion temperature in the PECVD process, CNTs can have larger diameters because of faster growth kinetics and enhanced carbon diffusion. At the interface between the CNTs and the active material (e.g., buffer layer and ITO), larger-diameter CNTs can provide a weaker electric field. This results in lower voltage output, which is less effective in inducing separation of photogenerated electron–hole pairs. This results in low current, which by Ohm's law is associated with decreased voltage output.

Because of changes induced by the plasma treatment, the effective surface area of the CNTs can decrease while the radius increases. This decrease in the active region available for charge generation and photon absorption results in lower current and voltage production.

2.2.2.4 The effect of absorber layer thickness (*t*) on the solar cell performance Figure 10 shows the variation in the solar cell efficiency with absorber layer thickness (t). It can be seen that when we increase the absorber layer thickness, the power output and efficiency of the solar cell increase. This is because an increase in the thickness of the absorber layer results in an increase in the number of CNTs, which leads to increased output voltage and thus an increase in the power output and efficiency of the solar cell [52, 53].

3 Comparisons of the efficiency of the simulated and analytical model of the proposed ITO-/CDS-/CNT-based solar cell

Figure 11 shows the efficiency of the constructed solar cell captured from simulation and analytical work. It can be seen that when we increase the value of t, the efficiency of the solar cell increases, and the slight difference between the analytical and simulation results may be the result of environmental effects such as differences in operating temperature and solar irradiance fluctuation.

4 Percentage error in solar cell efficiency between simulation and analytical work

The percentage error for the efficiency of the solar cell can be expressed as follows:

$$\text{Percentage error for the efficiency of the solar cell}(\%) = \left|\frac{\text{simulated value} - \text{analytical value}}{\text{simulated value}}\right| \times 100$$

Thus, the percentage error between simulation and analytical work is 1.71%. The results of calculations are shown in detail in Table 7.

5 Summary of results of the current work

A synopsis of the entire work is presented in Table 8.

6 Conclusion

In this work, PECVD was used to grow CNTs, which were incorporated into a solar cell as an absorber layer. Plasma parameters including electron and ion temperature and density as the result of PECVD treatment were then investigated with respect to their effects on the solar cell efficiency. Both simulation and theoretical modeling were carried out, and the results were very similar. This can be seen in the error calculation, which was about 1.7%. Here, we varied a number of plasma parameters that could impact solar cell efficiency. Since no previous studies have reported an expression that can directly relate plasma variables to solar cell efficiency, we solved this problem in the present work and successfully established a correlation between solar cell efficiency and plasma parameters. The results obtained from our work match very closely with those of recent research [25, 54–70]. The incorporation of CNTs and the efficiency achieved with respect to plasma parameters represents an important result and provides a foundation for further research. Here, we have used PECVD-grown CNTs in the absorber layer. For future work, CNTs can also be used in the hole transport layer (HTL), ITO-like layer, and electron transport layer (ETL), through which solar cell performance may be enhanced.

Acknowledgements The authors Suraj Kumar Singh and Ishu Sharma express their gratitude to Mr. Harender Mor for his valuable counsel.

Author contributions Contributions from all authors are equal.

Funding There were no grants, funds, or any other kinds of support granted.

Data availability The corresponding author is able to provide access to the datasets acquired and/or analyzed during this work upon justifiable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Salameh, Z.: Photovoltaic. In: Renewable energy system design, pp. 33–113. Elsevier, Amsterdam (2014). https://doi.org/10.1016/B978-0-12-374991-8.00002-7
2. Dambhare, M.V., Butey, B., Moharil, S.V.: Solar photovoltaic technology: a review of different types of solar cells and its future trends. J. Phys. Conf. Ser. (2021). https://doi.org/10.1088/1742-6596/1913/1/012053
3. Kumar K. (n.d.). A history of the solar cell, in patents. (2020). https://wiki-solar.org/library/public/190314_Utility-scale_solar_in_2018.pdf
4. Li, F., Liu, K., Dai, J.: Flexible p-i-n perovskite solar cell with optimized performance by KBF_4 additive. Opt. Express **32**, 366–378 (2024). https://doi.org/10.1364/oe.503856
5. Fraas, L., Partain, L.: Solar cells and their applications, 2nd Edition (Wiley Series in Microwave and Optical Engineering) (2010). https://www.researchgate.net/file.PostFileLoader.html?id=545b4859d685cc7d568b4651&assetKey=AS:273631823761408@1442250320282
6. Kirchartz, T., Rau, U.: What makes a good solar cell? Adv. Energy Mater. (2018). https://doi.org/10.1002/aenm.201703385
7. Chang, H.W., Li, C.C., Haung, T.Y., Li, H.C., Tan, C.S.: Enhancing perovskite solar cell efficiency and stability through the incorporation of BDAI2 and DMPDAI2. Energy Adv. (2024). https://doi.org/10.1039/d3ya00496a
8. Moshksar, E., Ghanbari, T.: Real-time estimation of solar irradiance and module temperature from maximum power point condition. IET Sci. Meas. Technol. **12**(6), 807–815 (2018). https://doi.org/10.1049/iet-smt.2017.0476
9. Khan, M.I., Fatima, N., Mustafa, G.M., Sabir, M., Abubshait, S.A., Abubshait, H.A., Alshahrani, T., Iqbal, M., Laref, A., Baig, M.R.: Improved photovoltaic properties of dye sensitized solar cell by irradiations of Ni2+ions on Ag-doped TiO_2 photoanode. Int. J. Energy Res. **45**(6), 9685–9693 (2021). https://doi.org/10.1002/er.6469
10. Wang, F., Matsuda, K.: Applications of carbon nanotubes in solar cells. In: Nakashima, N. (ed.) Nanocarbons for energy conversion: supramolecular approaches, pp. 497–536. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-319-92917-0_20
11. Iijima, S.: Helical microtubules of graphitic carbon. Nature **354**, 56–58 (1991). https://doi.org/10.1038/354056a0
12. Kim, M., Goerzen, D., Jena, P., Zeng, J., Meidl, R., Heller, D.: Human and environmental safety of carbon nanotubes across their life cycle. Nat. Rev. Mater. (2023). https://doi.org/10.26434/chemrxiv-2023-psth7
13. Dresselhaus, M.S., Dresselhaus, G., Saito, R.: Physics of carbon nanotubes. Carbon **33**(7), 883–891 (1995). https://doi.org/10.1016/0008-6223(95)00017-8
14. Takakura, A., Beppu, K., Nishihara, T., Fukui, A., Kozeki, T., Namazu, T., Miyauchi, Y., Itami, K.: Strength of carbon nanotubes depends on their chemical structures. Nat. Commun. (2019). https://doi.org/10.1038/s41467-019-10959-7
15. Attia, A., Saddam Hussain, M.I., Khan, A.S., Seliem, A.F., Mohammed, A.Y.A., Ibrahim, M.M.: Tuning the band gap edges of perovskite material by Cd doping for achieving high current density in perovskite solar cells. Ceram. Int. **49**(12), 20465–20469 (2023). https://doi.org/10.1016/j.ceramint.2023.03.176
16. Venkataraman, A., Amadi, E.V., Chen, Y., Papadopoulos, C.: Carbon nanotube assembly and integration for applications. Nanoscale Res. Lett. (2019). https://doi.org/10.1186/s11671-019-3046-3
17. Wilder, J.W.G., Venema, L.C., Rinzler, A.G., Smalley, R.E., Dekker, C.: Electronic structure of atomically resolved carbon nanotubes. Nature **391**, 59–62 (1998). https://doi.org/10.1038/34139
18. Iijima, S., Ichihashi, T.: Single-shell carbon nanotubes of 1-nm diameter. Nature **1993**(363), 603–605 (1993). https://doi.org/10.1038/363603a0
19. Bethune, D.S., Kiang, C.H., de Vries, M.S., Gorman, G., Savoy, R., Vazquez, J., Beyers, R.: Cobalt-catalysed growth of carbon nanotubes with single-atomic-layer walls. Nature **363**, 605–607 (1993). https://doi.org/10.1038/363605a0
20. Sugai, T., Yoshida, H., Shimada, T., Okazaki, T., Shinohara, H., Bandow, S.: New synthesis of high-quality double-walled carbon nanotubes by high-temperature pulsed arc discharge. Nano Lett. (2003). https://doi.org/10.1021/nl034183
21 Bandow, S., Takizawa, M., Hirahara, K., Yudasaka, M., Iijima, S.: Raman scattering study of double-wall carbon nanotubes derived from the chains of fullerenes in single-wall carbon nanotubes. Chem. Phys. Lett. (2001). https://doi.org/10.1016/S0009-2614(01)00192-0
22. Kumar, M.: 'Carbon Nanotube Synthesis and Growth Mechanism', carbon nanotubes - Synthesis, Characterization, Applications. InTech (2011). https://www.intechopen.com/chapters/16802
23 Sharma, P., Pavelyev, V., Kumar, S., Prabhash Mishra, S.S., Islam, N.T.: Analysis on the synthesis of vertically aligned carbon nanotubes: growth mechanism and techniques. J. Mater. Sci. Mater. Electron. **31**(6), 4399–4443 (2020). https://doi.org/10.1007/s10854-020-03021-6
24 Zhou, H.P., Xu, S., Xiao, S.Q.: High-density plasma-enhanced chemical vapor deposition of si-based materials for solar cell applications. In: Chemical Vapor Deposition - Recent Advances and Applications in Optical, Solar Cells and Solid State Devices. InTech (2016). https://doi.org/10.5772/63529
25. Oublal, E., Abdelkadir, A.A., Sahal, M.: High performance of a new solar cell based on carbon nanotubes with CBTS compound as BSF using SCAPS-1D software. J. Nanopart. Res. (2022). https://doi.org/10.1007/s11051-022-05580-7
26. Caglar, B., Bertran, S., Enric Jover, E.: Production of carbon nanotubes by PECVD and their applications to supercapacitors. Màster Oficial - Nanociència i Nanotecnologia (2010). http://hdl.handle.net/2445/11122
27. Xosrovashvili, G., Gorji, N.E.: Numerical simulation of carbon nanotubes/GaAs hybrid PV devices with AMPS-1D. Int. J. Photoenergy (2014). https://doi.org/10.1155/2014/784857
28. Ahmmed, S., Aktar, A., Hossain, J., Ismail, A.B.M.: Enhancing the open circuit voltage of the SnS based heter- ojunction solar cell using NiO HTL. Sol. Energy **207**, 693–702 (2020). https://doi.org/10.1016/j.solener.2020.07.003
29. Erkan, M.E., Chawla, V., Scarpulla, M.A.: Reduced defect density at the CZTSSe/CdS interface by atomic layer deposition of

Al_2O_3. J. Appl. Phys. (2016). https://doi.org/10.1063/1.4948947
30. SCAPS version 3.3.10 ("SCAPS3310") of April 2021, and subsequent versions Copyright:the University of Gent ("Belgium") Department ELIS (Electronics and Information Systems)Marc Burgelman – Koen Decock - Johan Verschraegen - Stefaan Degrave - Alex Niemegeers (SCAPS is developed under LabWindows of National Instruments) Contact:Marc.Burgelman@ugent.be
31. Burgelman, M., Nollet, P., Degrave, S.: Modelling poly- crystalline semiconductor solar cells. Thin Solid Films **361**, 527–532 (2000). https://doi.org/10.1016/S0040-6090(99)00825-1
32 Markvart, T., Castañer, L.: Semiconductor materials and modeling. In: McEvoy's handbook of photovoltaics: fundamentals and applications, pp. 29–57. Elsevier, Amsterdam (2018). https://doi.org/10.1016/B978-0-12-809921-6.00002-1
33. Mellit, A., Kalogirou, S.A.: A survey on the application of artificial intelligence techniques for photovoltaic systems. In: McEvoy's handbook of photovoltaics, pp. 735–761. Academic Press, Cambridge (2018). https://doi.org/10.1016/B978-0-12-809921-6.00019-7
34. Work-function values. (n.d.). https://public.wsu.edu/~pchemlab/documents/Work-functionvalues.pdf
35 Khorasani, A.E., Schroder, D.K., Alford, T.L.: Optically excited MOS-capacitor for recombination lifetime measurement. IEEE Electron Device Lett. **35**(10), 986–988 (2014). https://doi.org/10.1109/LED.2014.2345058
36 Pelant, I., Valenta, J.: Luminescence of disordered semiconductors. In: Pelant, I., Valenta, J. (eds.) Luminescence spectroscopy of semiconductors, pp. 242–262. Oxford University Press, Oxford (2012). https://doi.org/10.1093/acprof:oso/9780199588336.003.0009
37 Stranks, S.D., Burlakov, V.M., Leijtens, T., Ball, J.M., Goriely, A., Snaith, H.J.: Recombination kinetics in organic-inorganic perovskites: excitons, free charge, and subgap states. Phys. Rev. Appl. (2014). https://doi.org/10.1103/PhysRevApplied.2.034007
38 Li, S.S. (ed.): Semiconductor Physical Electronics. Springer New York, New York, NY (2006). https://doi.org/10.1007/0-387-37766-2
39. Nisoli, M.: Semiconductor photonics. Societa editrice Esculapio (2016). ISBN 978–8893850025. OCLC 964380194.
40 Svelto, O.: Types of lasers. In: Svelto, O. (ed.) Principles of Lasers, pp. 287–377. Springer US, Boston, MA (1999). https://doi.org/10.1007/978-1-4615-7670-9_6
41. Blumenau, A.T.: Dislocation related photoluminescence in silicon. Phys. Rev. Lett. **87**(18), 187404 (2001). https://doi.org/10.1103/PhysRevLett.87.187404.18
42 van Dijken, A., Meulenkamp, E.A., Vanmaekelbergh, D., Meijerink, A.: The kinetics of the radiative and nonradiative processes in nanocrystalline ZnO particles upon photoexcitation. J. Phys. Chem. B **104**(8), 1715–1723 (2000). https://doi.org/10.1021/jp993327z
43 Shockley, W., Read, W.T.: Statistics of the recombinations of holes and electrons. Phys. Rev. **87**(5), 835–842 (1952). https://doi.org/10.1103/PhysRev.87.835
44. Hall, R.N.: Germanium rectifier characteristics. Phys. Rev. **83**(1), 228 (1951)
45 Kandada, A.R.S., D'Innocenzo, V., Lanzani, G., Petrozza, A.: Photophysics of Hybrid Perovskites. In: Unconventional Thin Film Photovoltaics, pp. 107–140. The Royal Society of Chemistry, London (2016). https://doi.org/10.1039/9781782624066-00107
46. Stevenson, R.: The LED's Dark Secret: Solid-state lighting won't supplant the lightbulb until it can overcome the mysterious malady known as droop. IEEE Spectrum (August 2009).
47. Justin, I., Lucio, M., Jacques, P., Speck, J.S.: Claude Weisbuch. Cause of LED efficiency droop finally revealed. Phys. Rev. Lett. (2013).
48. Nelson, J.: The physics of solar cells, p. 116. Imperial College Press, London (2003)
49. Eades, W.D., Swanson, R.M.: Calculation of surface generation and recombination velocities at the Si-SiO2 interface. J. Appl. Phys. **58**(11), 4267–4276 (1985). https://doi.org/10.1063/1.335562.ISSN0021-8979
50. Recombination in low mobility semiconductors: Langevin theory". 4 April 2008.
51. Lakhwani, G., Rao, A., Friend, R.H.: Bimolecular recombination in organic photovoltaics. Annu. Rev. Phys. Chem. **65**(1), 557–581 (2014). https://doi.org/10.1146/annurev-physchem-040513-103615
52 Kansal, M., Sharma, S.C.: Plasma-based nanoarchitectonics for vertically aligned dual-metal carbon nanotube field-effect transistor (VA-DMCNFET) device: effect of plasma parameters on transistor properties. Appl. Phys. A Mater. Sci. Process. (2022). https://doi.org/10.1007/s00339-021-05096-2
53. Sharma, S.C., Tewari, A.: Effect of plasma parameters on growth and field emission of electrons from cylindrical metallic carbon nanotube surfaces. Phys. Plasmas (2011). https://doi.org/10.1063/1.3610375
54 Mahammedi, N.A., Gueffaf, H., Lagoun, B., Ferhat, M.: Numerical simulation and optimization of a silicon clathrate-based solar cell n-Si136/p-Si2 using SCAPS-1D program. Opt. Mater. (2020). https://doi.org/10.1016/j.optmat.2020.110043
55. Naureen, S., Lohia, P., Dwivedi, D., Ameen, S.: A comparative study of quantum dot solar cell with two different ETLs of WS2 and IGZO using SCAPS-1D simulator. Solar **2**(3), 341–353 (2022). https://doi.org/10.3390/solar2030020
56. Kartopu, G., Williams, B.L., Zardetto, V., Gürlek, A.K., Clayton, A.J., Jones, S., Kessels, W.M., Creatore, M., Irvine, S.J.C.: Enhancement of the photocurrent and efficiency of CdTe solar cells suppressing the front contact reflection using a highly-resistive ZnO buffer layer. Sol. Energy Mater. Sol. Cells **191**, 78–82 (2019). https://doi.org/10.1016/j.solmat.2018.11.002
57 Mizutani, T., Ohnaka, H., Okigawa, Y., Kishimoto, S., Ohno, Y.: A study of preferential growth of carbon nanotubes with semiconducting behavior grown by plasma-enhanced chemical vapor deposition. J. Appl. Phys. (2009). https://doi.org/10.1063/1.3234389
58 Abdelaziz, S., Zekry, A., Shaker, A., Abouelatta, M.: Investigating the performance of formamidinium tin-based perovskite solar cell by SCAPS device simulation. Opt. Mater. (2020). https://doi.org/10.1016/j.optmat.2020.109738
59 Fossum, J.G., Mertens, R.P., Lee, D.S., Nijs, J.F.: Carrier recombination and lifetime in highly doped Silicont. In Solid-State Electron. (1983). https://doi.org/10.1016/0038-1101(83)90173-9
60 Seck, S.M., Ndiaye, E.N., Fall, M., Charvet, S.: Study of efficiencies CdTe/CdS photovoltaic solar cell according to electrical properties by scaps simulation. Nat. Resour. **11**(04), 147–155 (2020). https://doi.org/10.4236/nr.2020.114009
61 Wang, F., Kozawa, D., Miyauchi, Y., Hiraoka, K., Mouri, S., Ohno, Y., Matsuda, K.: Considerably improved photovoltaic performance of carbon nanotube-based solar cells using metal oxide layers. Nat. Commun. (2015). https://doi.org/10.1038/ncomms7305
62 Tifidat, K., Maouhoub, N., Benahmida, A.: An efficient numerical method and new analytical model for the prediction of the five parameters of photovoltaic generators under non-STC conditions. E3S Web Conf. **297**, 01034 (2021). https://doi.org/10.1051/e3sconf/202129701034
63 Markose, K.K., Jasna, M., Subha, P.P., Antony, A., Jayaraj, M.K.: Performance enhancement of organic/Si solar cell using CNT

embedded hole selective layer. Sol. Energy **211**, 158–166 (2020). https://doi.org/10.1016/j.solener.2020.09.024
64. Shihab Uddin, M., Khalid Hossain, M., Uddin, M.B., Toki, G.F.I., Ouladsmane, M., Rubel, M.H.K., Tishkevich, D.I., Sasikumar, P., Haldhar, R., Pandey, R.: An in-depth investigation of the combined optoelectronic and photovoltaic properties of lead-free Cs2AgBiBr6 double perovskite solar cells using DFT and SCAPS-1D frameworks. Adv. Electron. Mater. (2024). https://doi.org/10.1002/aelm.202300751
65. Lee, C., Lee, S.W., Bae, S., Shawky, A., Devaraj, V., Anisimov, A., Kauppinen, E.I., Jin-Woo, O., Kang, Y., Kim, D., Jeon, Il., Maruyama, S., Lee, H.S.: Carbon nanotube electrode-based perovskite-silicon tandem solar cells. Solar RRL (2020). https://doi.org/10.1002/solr.202000353
66 Liu, S.H., Yuan, J.R., Wu, Y., Deng, X.H., Yu, Q.M.: Numerical simulation of Sb2Se3-based solar cells. Chalcogenide Lett. **21**(3), 229–241 (2024). https://doi.org/10.15251/CL.2024.213.229
67 Wieland, L., Li, H., Rust, C., Chen, J., Flavel, B.S.: carbon nanotubes for photovoltaics: from lab to industry. Adv. Energy Mater. **11**, 2002880 (2021). https://doi.org/10.1002/aenm.202002880
68. Najim, A., Bajjou, O., Bakour, A., Moulaoui, L., Rahmani, K.: Numerical computing of CdTe-based solar cells with graphene oxide buffer layers using SCAPS-1D software. Indian J. Phys. **98**(1), 67–77 (2024). https://doi.org/10.1007/s12648-023-02774-0
69. Benali, H., Hartiti, B., Lmai, F., Batan, A., Fadili, S., Thevenin, P.: Enhancing the efficiency of the organic-inorganic hybrid perovskite cells using Al-doped ZnO as an electron transport layer and CNTs as a hole transport layer: an experimental and numerical study. Optik (2024). https://doi.org/10.1016/j.ijleo.2024.171733
70. Forhad Hossain, M., Ghosh, A., Mamun, M.A.A., Miazee, A.A., Hamad Al-lohedan, R., Ramalingam, J., Buian, M.F.I., Karim, S.R.I., Yousup Ali, M., Sundararajan, M.: Design and simulation numerically with performance enhancement of extremely efficient Sb2Se3-Based solar cell with V2O5 as the hole transport layer, using SCAPS-1D simulation program. Opt. Commun. (2024). https://doi.org/10.1016/j.optcom.2024.130410

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Cardiovascular Disease with Machine Learning: A Comparative Analysis of Classification Models

Diya Malhotra

diyamalhotra014@gmail.com

Delhi Technological University Department of Applied Mathematics https://orcid.org/0009-0007-1249-1168

Aarushi Saini

Delhi Technological University Department of Applied Mathematics

Research Article

Keywords: Cardiovascular Diseases, Machine learning, Logistic Regression, SVM, KNN, Decision Tree, Random Forest, Naive Bayes, Roc-Auc

Posted Date: July 22nd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4297363/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Predicting Cardiovascular Disease with Machine Learning: A Comparative Analysis of Classification Models

Diya Malhotra[1], Aarushi Saini[2]
[1]Department of Applied Mathematics, Delhi Technological University (DTU), Delhi, India

[2]Department of Applied Mathematics, Delhi Technological University (DTU), Delhi, India

diyamalhotra014@gmail.com

aarushisaini007@gmail.com

Under the supervision of Prof. Anjana Gupta (DTU)

Abstract:

Cardiovascular disease is a very serious health issue. So, in order to prevent its spread, we need to understand the reason behind it's increase. Different factors like our lifestyle, our genes, the surroundings that we live in and so on all contribute to the risk of getting CVD. So, it is important to make positive changes in our day-to-day life which in the end will make us healthier. This research paper delves into understanding the importance behind such factors. We have used various classification models such as Logistic Regression, Decision Tree Algorithm, Random Forest, KNN, Support vector machine and Naïve Bayes to make predictions regarding cardiovascular disease patients. We have used data from UCI Repository that includes the features (predicator variables) such as age, BMI, gender, cholesterol, alcohol intake and so on to determine the presence of cardiovascular disease patients (response variable). Different models have been used to find out which model works best and we have done this by estimating various metrics that are essential for the assessment of model performance such as accuracy, precision, recall, etc. The Support Vector Machine model had the highest accuracy, Roc-Auc. So, this shows that the Support Vector Machine (SVM) so far is the best model for making predictions regarding CVD.

Keywords: Cardiovascular Diseases; Machine learning; Logistic Regression; SVM; KNN; Decision Tree; Random Forest; Naive Bayes; Roc-Auc.

Statement and Declaration:

Competing Interests and Funding:

We assert no competing interests concerning the research presented herein. The study was conducted independently and self-funded.

Author Contributions:

Aarushi Saini and Diya Malhotra played important roles in conducting this research including methodology, analysis, manuscript composition under the supervision of Prof. Anjana Gupta.

Ethical Approval:

This study was conducted in strict adherence and received approval from institution review board.

1.Introduction

The heart is a vital organ that is present in both human and animal bodies and is crucial for the survival. It circulates blood throughout the body via the circulatory/vascular system. Cardiovascular disease refers to a group of diseases that affect your heart and blood vessels. These diseases can impact your heart or blood vessels in one or more ways. A person might be symptomatic (feeling the disease physically) or asymptomatic (not feeling anything at all).

According to the World Health Organization (WHO) statistics, cardiovascular diseases (CVDs) were responsible for the highest number of deaths in 2019, representing 32% of all global deaths. The majority of the fatalities around 85% resulted from heart attacks and strokes.[1]. Populations that are growing older and becoming more numerous are making matters more complicated; in 2030, 22.2 million deaths from CVD are predicted [2]. Around 75% of cardiovascular fatalities happen in low and middle income nations. Heart disease is caused due to various risk factors, including physical inactivity, unhealthy diet, and the effective use of alcohol and tobacco. Most cardiovascular diseases can be averted by adopting a healthy daily lifestyle, which includes limiting salt in the diet, eating more vegetables and fruits, engaging in regular physical activity, and quitting smoking and excessive alcohol consumption [3]. Early detection of cardiovascular illness is crucial for the initial start of medication and counselling-based therapy.

Currently, computers are being used in various domains including medicine where they enhance decision-support system. Machine Learning (ML) is becoming more used in a variety of disciplines, including disease diagnosis in health care. Many researchers and healthcare professionals demonstrate the prospective of machine learning-based disease diagnostics (MLBDD), which is both affordable and time-efficient [4]. Health care data, such as pictures (X-ray, MRI) and tabular data (patient conditions, age, and gender), are utilised in developing MLBDD systems. ML is an instance of artificial intelligence (AI) that is becoming increasingly popular in the field of cardiovascular health. It is simply how computers process data and decide or classify tasks with or without human intervention. The ML conceptual framework is built on models that receive input data (e.g., images or text) and predict outcomes (e.g.,

favourable, unfavourable, or neutral) using a combination of mathematical optimization and statistical analysis. Several ML algorithms have been applied to everyday activities. Numerous causes of heart disease can complicate prediction.

With the help of machine learning techniques, heart disease can be identified. Various techniques are used to reliably and precisely predict heart disease. Feature selection methods such as Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), Naive Bayes (NB), and Support Vector Machine (SVM) and many more have been developed to be equally effective in disease prediction. We aim to contribute to the development of more effective and efficient heart disease prediction models by investigating various methodologies.

The remaining section of this paper is organised as follows: Section II describes the approach in depth, including data collection, data preparation, and the use of logistic regression, random forest, decision tree, SVM, and Naïve Bayes classifiers. Section III compares the metrics of different classifiers using experimental results and performance evaluation indicators. Finally, Section IV summarises the important findings and their significance in the field of heart disease prediction.

2.Research Methodology

Data: The aim of the research paper is to predict the people suffering from cardiovascular disease. For that we have taken the dataset regarding cardiovascular disease from UCI Repository and it consists of 70000 records, each representing a patient's information. This dataset includes data on observable traits and risk factors related to heart attacks. Table 1 in the Appendix provides an overview of the dataset.

Data Preprocessing: After the data has been collected, Preprocessing was done that involves dealing with missing values, correcting incorrect records, inconsistencies and so on. We made use of some statistical methods such as mean to fill in any missing information and also removed duplicated values. This was done to ensure that our research is based on complete and accurate dataset.

Feature Selection: Refers to a process where the most important features are selected from the original dataset and neglecting irrelevant ones. This process helps in improving the performance of model as too many features may lead to overfitting, complexity and so on.

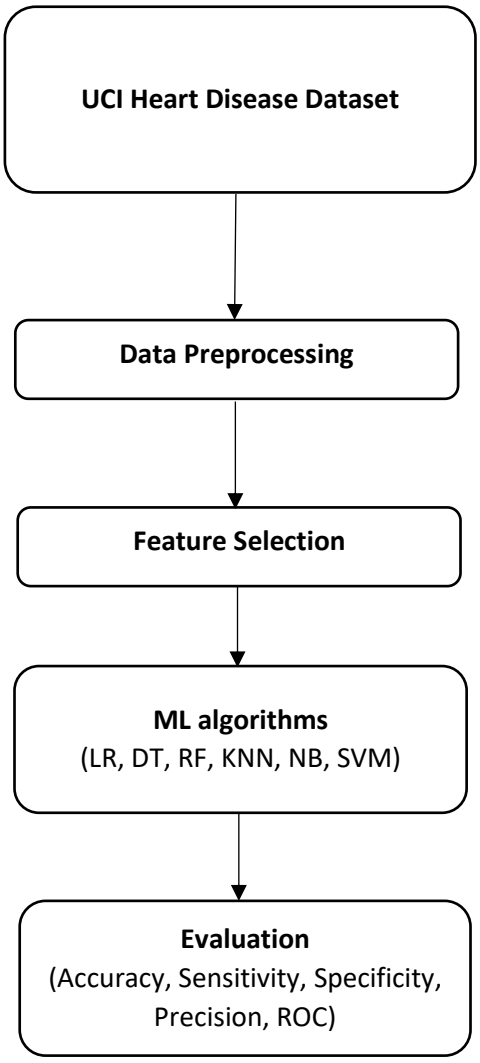


Figure 1. Research Methodology

2.1 Classification: Classification is the process of organizing a given dataset into distinct categories, a process which is applicable to both structured and unstructured data. The initial step in classification is predicting the category of the data points given. There are various methods in classification:

2.1.1 Logistic Regression: Logistic Regression (LR) is a supervised machine learning algorithm which aims at figuring out the chances of something happening based on different factors. It's like trying to predict if the person will have the cardiovascular disease or not. The LR technique is implemented in this study to classify heart disease. To boost performance, pre-processing of the data is performed, such as cleaning and detecting missing values. The most important part is feature selection, which improves algorithm accuracy and especially focuses on algorithm behaviour. Logistic regression's behaviour is that as training grows, so does prediction accuracy. The dependent variable in logistic regression is always binary. Logistic regression is mostly used for prediction and calculating the likelihood of success.

The sigmoid function is employed as a cost function to limit the logistic regression hypothesis between 0 and 1 (squashing), i.e., 0 h (x) 1 [5]. We classified and predicted the CVD patients in the machine learning LR using the function:

$$f(x) = \begin{cases} 1, & CVD\ present \\ 0, & CVD\ absent \end{cases}$$

In LR, the cost function is referred to as:

$$\log \frac{P(x)}{1 - P(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_P x_p \ldots .$$

2.1.2 Decision Tree: - The decision tree (DT) stands out as the most important predictive modelling and classification method within learning algorithms, extensively applied in practical approaches in supervised learning techniques [6, 7]. With the help of Decision Tree algorithm, a tree like structure is created which includes internal node (decisions made on the basis of features available), branches (outcomes of the decision) and leaf node (represents final decision).

While making predictions regarding cardiovascular disease through decision tree algorithm we first divided the dataset into training and testing datasets and then by calculating entropy which determines which feature to be used at various nodes. This process continues until a specific criterion is met i.e. all the decision nodes have been made.

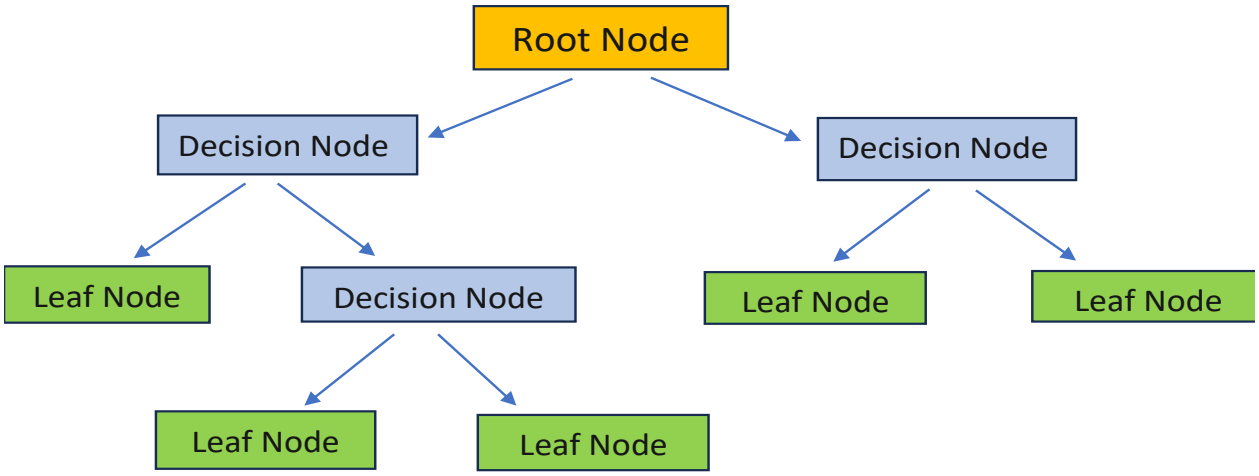


Figure 2: Decision Tree

2.1.3 Random Forest: The Random Forest (RF) is a supervised machine learning technique that uses decision trees to perform classification and prediction by averaging several independent base models [8]. The Random Forest creates a set of decision trees by randomly selecting a subset of the training set. It is essentially a set of decision trees (DT) drawn from a randomly chosen subset of the training set, and it then aggregates the votes from the many decision trees to determine the final prediction [9]. The greater the number of trees in the forest, the higher the accuracy and the lower the possibility of overfitting. The working of Random Forest algorithm is demonstrated in the below diagram [10]:

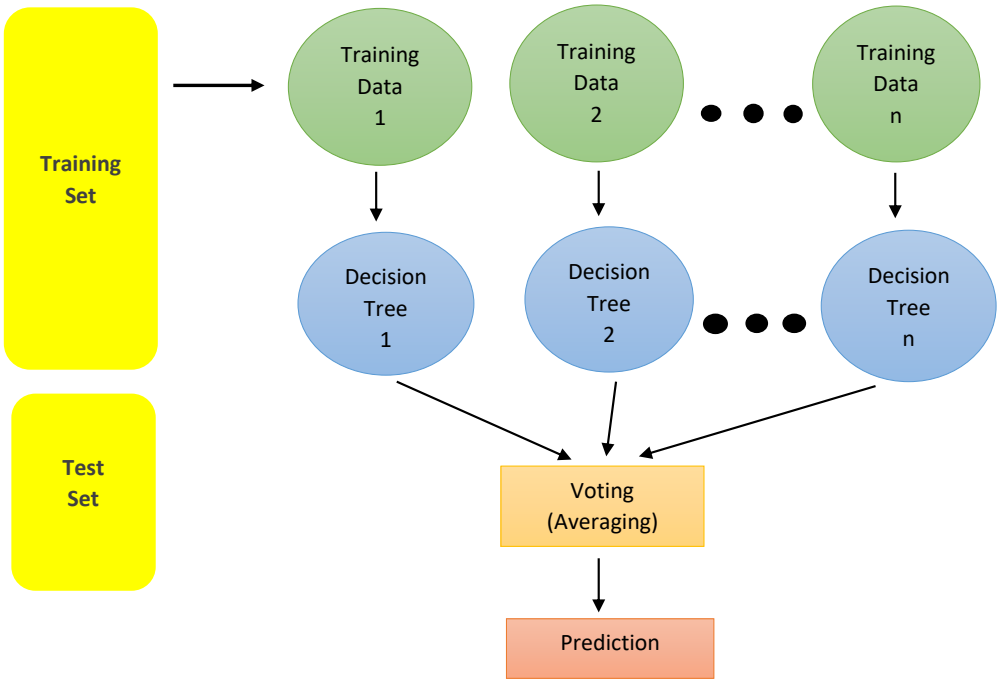


Figure 3: Random Forest algorithm

2.1.4 K-Nearest Neighbour (KNN): The K-Nearest Neighbours algorithm is a type of machine learning approach used for classification tasks. It works by determining the class of a given data point based on the distances between the point and the points in the training dataset. In KNN we need to find out the k neighbours in the space and then with the help of k neighbours we perform the classification i.e. KNN makes prediction on the basis of what majority of k neighbours indicate. For example, in case of predicting cardiovascular disease, if a lot of neighbours have the disease, then the new person also has the disease.

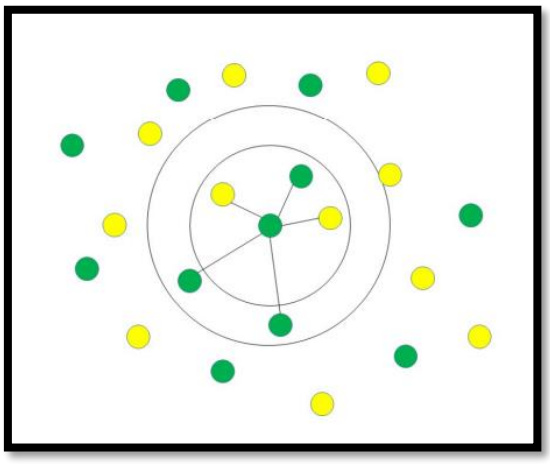


Figure 4: Demonstration of K-NN identifying unknown pattern by assigning a value to the K, where the nearest neighbour category of the K training sample is considered the same as the classification [11].

2.1.5 Support Vector Machine (SVM): SVM is commonly used for the purpose of classification. The aim of the SVM is to find a hyperplane that well separates the data points that belongs to different classes. SVM aims to maximize the margin which is the distance between the hyperplane and the nearest support vectors from each set. Larger the margin means there is gap between the classes and in turn we are able to draw the best possible line i.e. hyperplane that best separates the data as shown in the figure.

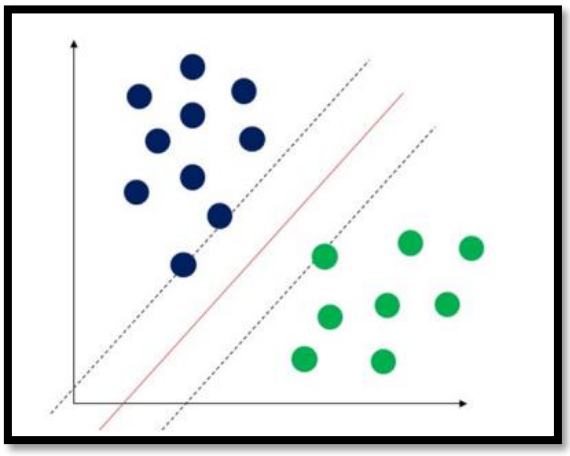


Figure 5: Demonstration of support vector machine. In the illustration, the solid red line represents the separating hyperplane, while the distance between two dotted lines depicts the maximum margin for separating distinct classes [12].

2.1.6 Naïve Bayes: The Naive Bayes algorithm is a supervised learning technique that uses the Bayes theorem with strong (Naïve) independence assumptions between features to solve problems regarding classification. It is a probabilistic classifier i.e. it predicts on the basis of an object's probability [13]. The Naïve Bayes classifier stands out as simple and efficient classification algorithm, hence making it a popular choice for quickly creating machine learning models making predictions swiftly.

Internal calculations will be performed to determine the probability of each value in relation to the feature. We take into consideration the value with the highest probability once we have computed the probability of each value of the output attributes [14,15]. Bayes algorithm is given by

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

where,

P(A|B) refers to Posterior probability: Probability of hypothesis A on the observed event B.

P(B|A) refers to Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

P(A) refers to Prior Probability: Denoting the likelihood of hypothesis before considering the evidence.

P(B) refers to Marginal Probability: Probability of Evidence.

2.2 Assessment of model performance

To find out whether the model is performing successfully and making right predictions on the basis of new unseen data we use metrics which are used for almost all machine learning algorithms.

These are: -

1) Confusion Matrix:

	Actual Positive	Actual Negative
Predicted Positive	True Positive	False Positive
Predictive Negative	False Negative	True Negative

True positive are the outcomes that are correctly predicted as positive. For example, in case of Cardiovascular disease, True positive finds out that the person tested with cardiovascular disease actually have the disease whereas True Negative finds out that the person tested do not have the disease and indeed they do not have disease.

False Positive: Outcomes that are actually negative but incorrectly predicted as positive. For example, the test says person have the disease but actually they do not have and False Negative is opposite of False Positive.

2) Accuracy:

Accuracy is the ratio of sum of total number of true positives and true negatives to total number of instances.

i.e.

$$Accuracy = \frac{\text{Total number of true positives } + \text{ Total number of false positives}}{\text{Total number of instances}}$$

for example, in case of predicting cardiovascular disease, how well the machine learning model correctly identify patients with or without disease. The more the accuracy means the model is able to make correct predictions most of the time.

3) Precision and Recall:

Precision is the ratio of total number of true positives to the sum of total number of true positives and total number of false positives.

i.e.

$$Precision = \frac{\text{Total number of true positives}}{\text{Total number of true positives } + \text{ Total number of false positives}}$$

Precision measures the positive outcome, for example, in the dataset taken above regarding cardiovascular diseases, it measures whether the person has disease. The higher the precision means the model is accurate meaning if the model is predicting that a person has cardiovascular disease, it is likely to be true.

Recall also known as Sensitivity or True positive rate represents the proportion of true positives among all actual positives. It is the total number of true positives to the sum of total number of true positives and total number of false negatives.

i.e.

$$Recall = \frac{\text{Total number of true positives}}{\text{Total number of true positives } + \text{ Total number of false negatives}}$$

For example, Recall identifies the number of patients who actually have the cardiovascular disease. Higher the recall means our ML model is able to identify all the persons who actually have cardiovascular disease.

4) Specificity

Specificity also known as true negative rate is the ratio of total number of true negatives to the sum of total number of true negatives and total number of false positives.

i.e.

$$Specificity = \frac{\text{Total number of true negatives}}{\text{Total number of true negatives } + \text{ Total number of false positives}}$$

For example, in case of predicting whether the patients have cardiovascular disease or not, specificity helps in finding out how well the model is predicting the patients that do not have disease.

5) Receiver Operating Characteristic (ROC)

The ROC curve is a probability graph that illustrates how well a classification model performs at various threshold values. It plots the true positive rate (TPR) against the false positive rate (FPR) across various classification thresholds.

TPR or true positive rate also known as Recall, can be calculated as:

$$TPR = \frac{TP}{TP + FN}$$

FPR or False Positive Rate can be calculated as:

$$FPR = \frac{FP}{FP + TN}$$

where,

TP: True Positive; **FP**: False Positive; **TN**: True Negative; **FN**: False Negative

3. Results and Discussion

We have used various classification models such as Logistic Regression, Decision Tree Algorithm, Random Forest, KNN, Support vector machine and Naïve Bayes to make predictions regarding cardiovascular disease patients. First, we commence by initiating the data cleaning process that involves removing null values, removing duplicates and filtering out irrelevant information. The following is done to ensure that our research is based on complete and accurate dataset.

Python code

1) Data Understanding

```
data=pd.read_csv("/content/cardio_train.csv",sep=";")
data.head()
```

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0

2) Data Preprocessing

i) Age Transformation, BMI Calculation, Column Removal

```
data['age_in_years']=(data['age']//365).astype('int')
data.drop(['id','age'],axis=1,inplace=True)
data['bmi']=data['weight']/(data['height']/100)**2
data.drop(['height','weight'],axis=1,inplace=True)
data.head()
```

	gender	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio	age_in_years	bmi
0	2	110	80	1	1	0	0	1	0	50	21.967120
1	1	140	90	3	1	0	0	1	1	55	34.927679
2	1	130	70	3	1	0	0	0	1	51	23.507805
3	2	150	100	1	1	0	0	1	1	48	28.710479
4	1	100	60	1	1	0	0	0	0	47	23.011177

ii) Dealing with null values and duplicated values

```
print(data.isnull().sum())
print("{} rows are same".format(data.duplicated().sum()))
data.drop_duplicates(inplace=True)
print("{} rows are same".format(data.duplicated().sum()))
```

```
gender          0
Systolic_BP     0
Dialostic_BP    0
cholesterol     0
glucose         0
smoker          0
alcohol         0
active          0
cardio          0
age_in_years    0
bmi             0
dtype: int64
3210 rows are same
0 rows are same
```

3) Training and Testing Data

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=Y, random_state=10)
print(X.shape, X_train.shape, X_test.shape, y_train.shape, y_test.shape)
```

```
(66790, 10) (53432, 10) (13358, 10) (53432,) (13358,)
```

i) Logistic Regression

```
# Logistic Regression
from sklearn.linear_model import LogisticRegression
log = LogisticRegression()
log.fit(X_train,y_train)
y_pred1 = log.predict(X_test)
```

ii) Decision Tree

```
[45] from sklearn import tree
     from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, classification_report
     # Initialize Decision Tree classifier
     clf = tree.DecisionTreeClassifier(max_depth=3, criterion='entropy')
     # Fit the classifier on the training data
     clf.fit(X_train, y_train)
     # Predictions on the test set
     y_pred = clf.predict(X_test)
```

iii) Random Forest

```
7] from sklearn.ensemble import RandomForestClassifier
   from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, f1_score, classification_report
   # Initialize RandomForestClassifier
   rf = RandomForestClassifier()
   # Fit the classifier on the training data
   rf.fit(X_train, y_train)
   # Predictions on the test set
   y_pred1 = rf.predict(X_test)
```

iv) KNN

```
from sklearn.preprocessing import StandardScaler
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, f1_score, classification_report
# Standardize the data
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
# Initialize KNN classifier with 12 neighbors
knn = KNeighborsClassifier(n_neighbors=12)
# Fit the classifier on the training data
knn.fit(X_train_scaled, y_train)
```

```
KNeighborsClassifier
KNeighborsClassifier(n_neighbors=12)
```

v) Naïve Bayes

```
[51] from sklearn.naive_bayes import GaussianNB
     from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, f1_score, classification_report
     # Initialize Gaussian Naive Bayes classifier
     nb = GaussianNB()
     # Fit the classifier on the training data
     nb.fit(X_train, y_train)
     # Predictions on the test set
     y_pred1 = nb.predict(X_test)
```

vi) SVM

```
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, confusion_matrix, classification_report
# Initialize Support Vector Machine (SVM) classifier
svm_classifier = SVC(random_state=50)
# Fit the classifier on the training data
svm_classifier.fit(X_train, y_train)
# Predictions on the test set
y_pred_svm = svm_classifier.predict(X_test)
```

We have also calculated various metrics that are essential for the assessment of model performance such as accuracy, precision, recall, etc. that can be seen from the figure below.

	Models	Accuracy	Precision	Recall	F1-score
0	Logistic Regression	65.990418	0.679325	0.635712	0.656795
1	Decision Tree	72.570744	0.753839	0.689237	0.720092
2	Random Forest	67.614912	0.678511	0.698157	0.688194
3	KNN	68.677946	0.722129	0.630886	0.673431
4	Naive Bayes	58.279683	0.735043	0.289266	0.415154
5	SVM	72.218895	0.770642	0.651068	0.705826

Figure 6: Assessment of model performance

From all the models we have found that the Decision Tree and Support Vector Machine have approximately same accuracy i.e. 72%. Higher Accuracy is important as it indicates the ability of model to make correct predictions. But accuracy alone might not be able to provide complete picture. Our main objective is to correctly identify the number of patients who actually have the cardiovascular disease.

So, sensitivity also known as recall is the metric that holds significant importance in the above context. Higher the sensitivity means our Machine Learning model is better at finding out larger proportion of persons who actually have the disease. We can also see the visualization of comparison of Model Metrics in Figure 7.

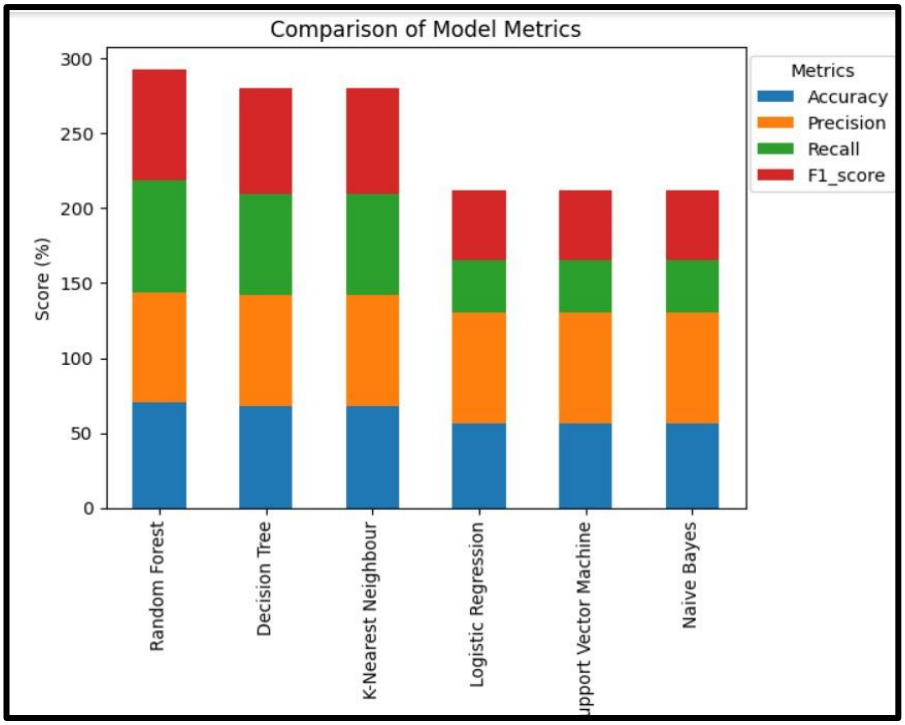


Figure 7: Visualization of metrics

Receiver operating characteristic (ROC) is also one of the important metrics that can be used to evaluate the performance of Machine Learning model i.e. ROC curve looks like a graph that can also show us what model is performing better. The Area under the curve (AUC) quantifies the overall effectiveness of the model, with a perfect classifier achieving an AUC of 1.0. Higher the score means that ROC-AUC is better at identifying the people who actually suffers from cardiovascular disease instead of making predictions that healthy people have the disease. In the analysis we have done, AUC (area under the curve) for Support Vector Machine (SVM) is highest i.e. at 79% value as illustrated in figure 8.

So, on the basis of our analysis, we have found that even though the Decision Tree model has slightly higher recall of 68% of class 1(with disease) whereas SVM achieves a slightly lower recall of 65% for the same class. But since our main objective is to correctly identify the number of patients who actually have the cardiovascular disease, it's important to consider the various performance metrics including ROC-AUC. Clearly from figure 8 we can see that SVM model outperforms all the machine learning models indicating that SVM is better at finding out larger proportion of persons who actually have the cardiovascular disease, making it a favourable choice for the data we have taken. Therefore, considering both accuracy and ROC-AUC, Support Vector Machine model appears to be more suitable for the dataset that we have taken.

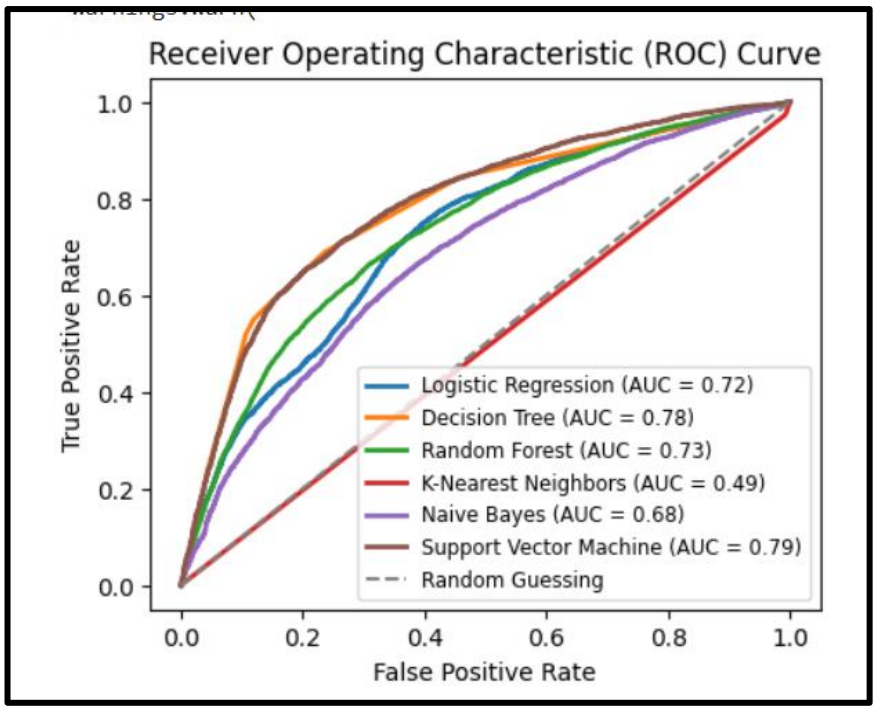


Figure 8: ROC-AUC

4. Conclusion

In medical data analysis, heart diseases are seen as a significant source of concern. Heart disease is a potentially fatal disease that affects millions of people worldwide each year (Asadi et al., 2021). As therefore, the early prediction of heart disease can benefit both patients and healthcare professionals by providing the statistical information required to avert deaths and expenditures. This research paper delves into the effectiveness of ML algorithms for predicting CVD patients. This research addressed the prediction problem of the UCI heart disease dataset including features such as age, BMI, gender, cholesterol, alcohol intake, and so on. The dataset consisted of labelled 70000 patients, it included both diagnosed heart disease patients and normal patients. To classify and predict CVD, six machine learning classification algorithms—DT, RF, LR, NB, KNN and SVM—were put into practice. It adhered to a suitable data science workflow, from data analysis and preparation to model creation, training, and evaluation. In order to identify the most appropriate machine learning algorithm within the model class, the performance of the selected ML method was estimated under a variety of conditions. The filter feature selection approach with the majority of features selected performed better than other methods in terms of the models Accuracy, Recall, Precision, and F-measures, according to the results.

The SVM algorithm had the highest accuracy of 72% for CVD. According to the above results, the SVM algorithm is the most suitable method for predicting cardiovascular disease patients for the dataset that we have considered. The ML algorithms used are restricted to research on heart disease prediction but we can use them for predicting a lot of other things. So, in general on the basis of our research we can see from ‘Figure 9’ that 56% of the people

are suffering from cardiovascular disease while 44% of the people are healthy according to the dataset that we have used.

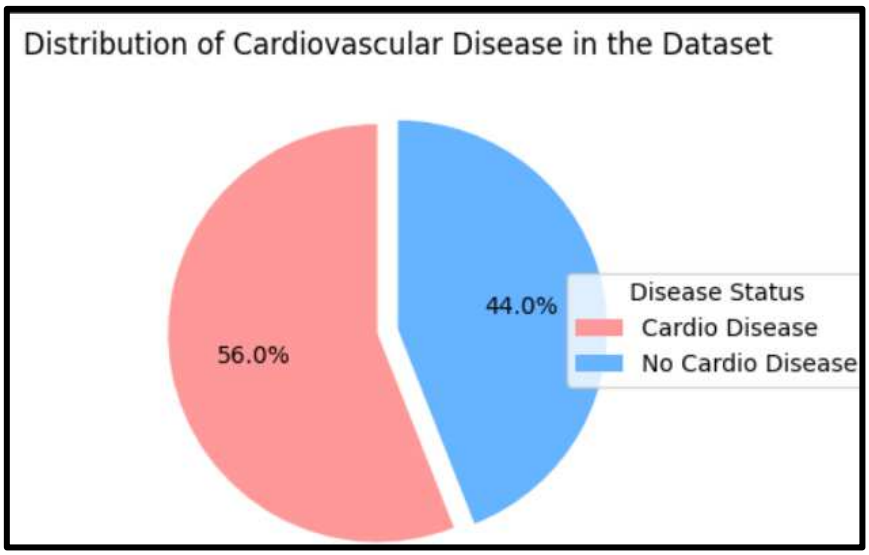


Figure 9: People having the disease and being healthy

Hence, we can conclude that machine learning has a significant role to play in our healthcare system. Historically, medical professionals used traditional processes and their intuition to diagnose diseases, which had drawbacks and resulted in higher expenses. However, with the use of machine learning models, diagnosis can be economically performed on large datasets.

5. References

[1] Y. Mamani-Ortiz, M. San Sebastián, A. X. Armaza et al., "Prevalence and determinants of cardiovascular disease risk factors using the WHO STEPS approach in Cochabamba, Bolivia," *Bmc Public Health*, vol. 19, no. 1, p. 786, 2019.

[2] G. A. Roth, M. H. Forouzanfar, A. E. Moran et al., "Demographic and epidemiologic drivers of global cardiovascular mortality," *New England Journal of Medicine*, vol. 372, no. 14, pp. 1333–1341, 2015.

[3] R. Ndejjo, G. Musinguzi, F. Nuwaha, H. Bastiaens, and R. K. Wanyenze, "Understanding factors influencing uptake of healthy lifestyle practices among adults following a community cardiovascular disease prevention programme in Mukono and Buikwe districts in Uganda: A qualitative study," PLoS One, vol. 17, no. 2, p. e0263867, Feb. 2022.

[4] Ahsan M.M., Siddique Z. Machine Learning-Based Heart Disease Diagnosis: A Systematic Literature Review.

[5] Kavya S M, PrathanyaSree C, Deepasindhu M, Nowshika B, Shijitha R, Heart Disease Prediction Using Logistic Regression, 30 December 2022, Journal of Coastal Life Medicine.

[6] T. G. Dietterich, “Machine learning,” Annual Review of Computer Science, vol. 4, no. 1, pp. 255–306, 1990.

[7] P. H. Swain and H. Hauska, “The decision tree classifier: design and potential,” IEEE Transactions on Geoscience Electronics, vol. 15, no. 3, pp. 142–147, 1977.

[8] Madhumita Pal and Smita Parija (2021), Prediction of Heart Diseases using Random Forest.

[9] Rajkumar Gangappa Nadakinamani,[1] A. Reyana,[2] Sandeep Kautish [3] A. S. Vibith, [4] Yogita Gupta, [5] Sayed F. Abdelwahab, [6] and Ali Wagdy Mohamed [7] (2022 Jan 11), Clinical Data Analysis for Prediction of Cardiovascular Disease Using Machine Learning Techniques.

[10] Sharath Pokala, Bandari Nithya, VOL12, ISSUE01,2021, “Random Forest Classifier based on Heart Disease Prediction”,Journal of Cardiovascular Disease Research, pp.799-801.

[11] Duneja, A.; Puyalnithi, T. Enhancing classification accuracy of k-nearest neighbours algorithm using gain ratio. Int. Res. J. Eng. Technol 2017, 4, 1385–1388.

[12] Yuan, R.; Li, Z.; Guan, X.; Xu, L. An SVM-based machine learning method for accurate internet traffic classification. Inf. Syst. Front. 2010, 12, 149–156.

[13] K. Vembandasamy[1], R. Sasipriya[2] and E. Deepa[3], Heart Diseases Detection Using Naive Bayes Algorithm, Vol. 2 Issue 9, September 2015.

[14] “A STUDY OF HEAR Y OF HEART DISEASE DI T DISEASE DIAGNOSIS USING M GNOSIS USING MACHINE LEARNING AND DATA MINING”, Intisar Ahmed, (12-2022), pp. 17-19.

[15] Uma N Dulhare, (May 21,2018), Prediction system for heart disease using Naive Bayes and particle swarm optimization

Appendix

Data features:

- Age | Objective Feature | age | int (years)
- Height | Objective Feature | height | int (cm) |
- Weight | Objective Feature | weight | float (kg) |
- BMI | Objective Feature | BMI | int |
- Gender | Objective Feature | gender | categorical code |
- Systolic blood pressure | Examination Feature | ap_hi | int |
- Diastolic blood pressure | Examination Feature | ap_lo | int |
- Cholesterol | Examination Feature | cholesterol | 1: normal, 2: above normal, 3: well above normal |
- Glucose | Examination Feature | gluc | 1: normal, 2: above normal, 3: well above normal |
- Smoking | Subjective Feature | smoke | binary |
- Alcohol | Subjective Feature | alco | binary |
- Physical Activity | Subjective Feature | active | binary |
- Presence or absence of cardiovascular disease | Target Variable | cardio | binary |

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
0	0	18393	2	168	62.0	110	80	1	1	0	0	1	0
1	1	20228	1	156	85.0	140	90	3	1	0	0	1	1
2	2	18857	1	165	64.0	130	70	3	1	0	0	0	1
3	3	17623	2	169	82.0	150	100	1	1	0	0	1	1
4	4	17474	1	156	56.0	100	60	1	1	0	0	0	0
5	8	21914	1	151	67.0	120	80	2	2	0	0	0	0
6	9	22113	1	157	93.0	130	80	3	1	0	0	1	0
7	12	22584	2	178	95.0	130	90	3	3	0	0	1	1
8	13	17668	1	158	71.0	110	70	1	1	0	0	1	0
9	14	19834	1	164	68.0	110	60	1	1	0	0	0	0
10	15	22530	1	169	80.0	120	80	1	1	0	0	1	0
11	16	18815	2	173	60.0	120	80	1	1	0	0	1	0
12	18	14791	2	165	60.0	120	80	1	1	0	0	0	0
13	21	19809	1	158	78.0	110	70	1	1	0	0	1	0
14	23	14532	2	181	95.0	130	90	1	1	1	1	1	0
15	24	16782	2	172	112.0	120	80	1	1	0	0	0	1

Table 1: Overview of the Dataset

Materials Science and Engineering B 307 (2024) 117536

Contents lists available at ScienceDirect

Materials Science & Engineering B

journal homepage: www.elsevier.com/locate/mseb

Predictive design of $KSnI_3$-based perovskite solar cells using SCAPS and machine learning model

Rahul Kundara *, Sarita Baghel

Department of Applied Physics, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
$KSnI_3$-based PSC
SCAPS-1D
Machine learning and HTLs

ABSTRACT

The commercial application of lead (Pb)-based Perovskite Solar Cells (PSCs) are restrained due to its toxicity. In this work, we have investigated potassium tin iodide ($KSnI_3$) which is a potential candidate for non-lead PSC. The SCAPS-1D simulation is employed to examine the effect of several Hole Transport Layers (HTLs) as follow: Spiro-OMeTAD, CuI, CuSCN, Cu_2O, PEDOT: PSS, CdTe, $MASnBr_3$, CFTS on the performance of $KSnI_3$-based PSCs. The impact of layer thickness, doping density (N_A) and defect density (N_t) has been analyzed. We have a trained Machine learning model to predict the performance matrices of PSCs. We have obtained a maximum efficiency of 23.08 % at the thickness of the $KSnI_3$ absorber layer of 550 nm and N_t of 10^{13} cm^{-3}. The ML model predicts the performance matrices of examined PSC with accuracy of ~ 89 %. Hence, the FTO/WS_2/$KSnI_3$/PEDOT: PSS/Au configuration is most suitable to manufacture highly efficient PV devices.

1. Introduction

Utilizing solar cells for solar energy harvesting presents a commendable approach mitigating concerns about global warming. The photovoltaic effect converts incident sunlight into electrical energy. In 2009, Perovskite solar cell (PSC) was reported with an efficiency of 3.8 % and has achieved maximum power conversion efficiency (PCE) of greater than 26 % till now [1]. It became popular among the scientific community due to its promising attributes such as tunable bandgap (E_g), high absorption coefficient, high carrier mobility, low manufacturing cost and high PCE or efficiency [2,3]. Perovskite material has an ABX_3 structure (where A is an organic or inorganic cation, B is an inorganic metal cation and X is halogen anion) [4]. Despite the high performance of lead (Pb) based PSC, its instability due to rapid oxidation of Pb cation and toxic nature restricted its commercialization [5]. There is a need for alternative perovskite materials that are both less toxic and environmentally friendly while also having the potential to deliver exhibit good photovoltaic (PV) outcomes [6]. Hence there have been various investigations to replace hazardous Pb ions with other alternatives such as Sn [7], Ge [8], Bi [9], Sb [10] and Ag [11]. The tin (Sn) is a suitable choice due to comparable ionic radii and the phenomenal semiconducting properties of $APbX_3$. The study comprises investigation of $KSnI_3$ as a perovskite material which is used as an absorber in a PSC. In past reported investigations there is very limited study on $KSnI_3$ PSC. Grishma et al. (2023) have reported efficiency of 9.77 % [12]. So there needs to be proper optimization to enhance the efficiency of PSC. Whereas, in this work we have obtained a maximum PCE of 23.08 %. The SCAPS enabled machine learning model is used to optimize and predict the best efficiency [13]. It has been exhibited that ML is helpful in forecasting the design and performance of PSCs [14]. Firstly, we have simulated $KSnI_3$-based PSC with several HTLs (Spiro-OMeTAD, CuI, CuSCN, Cu_2O, PEDOT: PSS, CdTe, $MASnBr_3$ and CFTS) and $KSnI_3$ is employed as a perovskite absorber layer. Secondly, we have optimized the effect of distinct performance parameters of PSCs such as absorber layer thickness, defect density (N_t) and doping density (N_A). Ultimately, we trained an ML model to predict the performance parameters of the PSCs with an accuracy of approximately 89 %. The present method serves as a sturdy foundation for subsequent research and investigations driven by artificial intelligence (AI). The efficient PSCs can be designed with application of ML modelling and SCAPS-1D assisted by previous studies.

2. Device architecture and methods

The PSC consists of $KSnI_3$ absorber material with layer architecture: FTO/ETL/Perovskite layer/HTL/Back-contact depicted in Fig. 1. All the input data used in SCAPS simulation obtained from experimental and theoretical work enclosed is in Table 1 for FTO/TiO_2/$KSnI_3$/Spiro-

* Corresponding author.
E-mail address: rahul_2k20phdap10@dtu.ac.in (R. Kundara).

https://doi.org/10.1016/j.mseb.2024.117536
Received 22 May 2024; Accepted 24 June 2024
Available online 28 June 2024
0921-5107/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

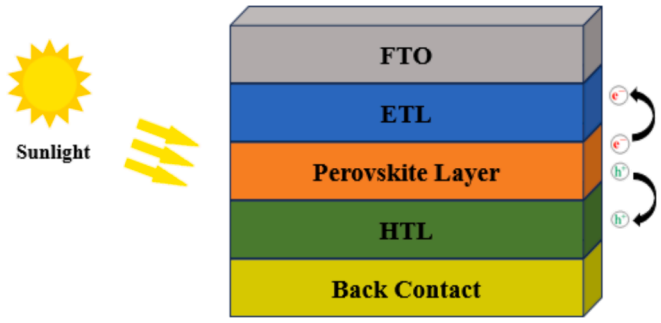


Fig. 1. Architecture of simulated PSCs.

Table 1
Input parameters of simulated PSC.

Parameter	FTO [15,16]	WS_2 [17]	$KSnI_3$ [12]	Spiro-OMeTAD [18,19]
Thickness (nm)	400	100*	300	50
Eg (eV)	3.5	1.8	1.84	3.0
χ (eV)	4.0	3.95	3.44	2.45
ε	9.0	13.6	10.4	3.0
N_C (cm^{-3})	2.02×10^{18}	1×10^{18}	2.2×10^{18}	1×10^{19}
N_V (cm^{-3})	1.8×10^{19}	2.4×10^{19}	1.8×10^{19}	1×10^{19}
μ_n (cm^2/Vs)	20	100	21.28	1×10^{-4}
μ_p (cm^2/Vs)	10	100	19.46	1×10^{-4}
N_D (cm^{-3})	2×10^{19}	1×10^{18}	1×10^{15}	0
N_A (cm^{-3})	0	0	1×10^{15}	2×10^{18}
N_t (cm^{-3})	1×10^{15}	$1 \times 10^{13*}$	1×10^{13}	1×10^{14}

OMeTAD/Au-Metal and HTLs in Table 2.

2.1. SCAPS-1D

The data of different parameters are chosen from previous studies. The SCAPS-1D (version 3.3.10) has been employed for simulation. This software can be used for simulations at different input physical conditions. This tool calculation is dependent on Poisson's and continuity equations. It can calculate Quantum efficiency (QE), J-V curve and carrier lifetime for both electrons (e^-s) and holes (h^+s).

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{q}{\epsilon}[p(x) - n(x) + N_D - N_A + \rho_p - \rho_n = 0 \quad (1)$$

$$\frac{1}{q}\frac{dJ_p}{dx} = G_{op}(x) - R(x) \quad (2)$$

$$\frac{1}{q}\frac{dJ_n}{dx} = -G_{op}(x) + R(x) \quad (3)$$

The simulations were run at 300 K using a "air mass 1.5 global" spectrum and a light power of 1000 W/m^2. The work is dependent on change in thickness, N_t, and N_A at which simulation of the proposed PSC has been done. The outcome of the device has been reflected by the I–V characteristics which represent the following PV parameters: J_{SC}, FF, V_{OC}, and PCE. The simulation procedure is depicted in Fig. 2.

2.2. Machine learning

Python was chosen as the programming language for implementing machine learning (ML) due to its widespread use in AI applications [27–29]. To work with the dataset, Numpy and Pandas libraries were utilized. The dataset pertains to solar cell features and was thoroughly cleaned and preprocessed to make it compatible with ML algorithms. Initially, a simple Random Forest algorithm from Scikit-Learn (Sklearn) was employed to determine the significance of each feature for the final model. As a result of this phase, eleven features were selected for further analysis. For feature extraction related to composition of perovskites, Pymatgen and Matminer packages were utilized. The data was then split

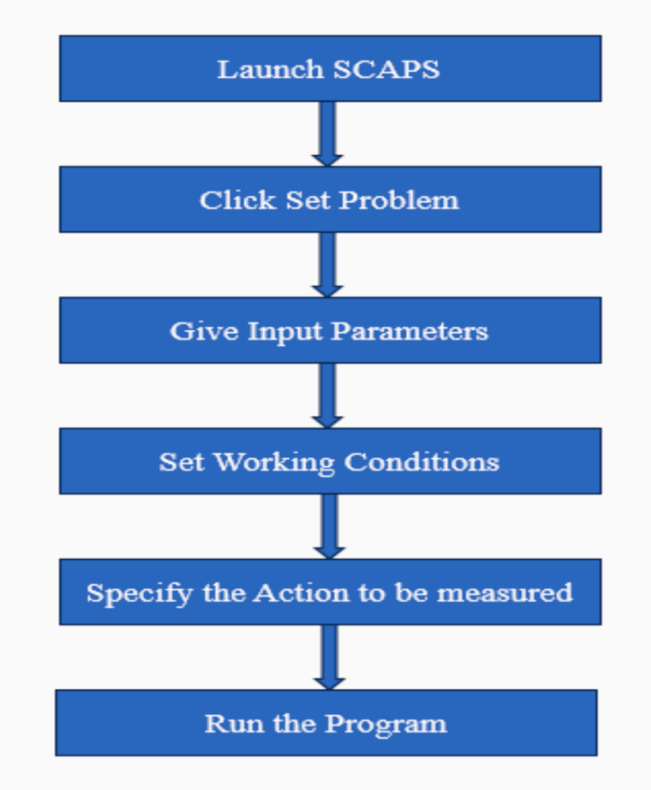


Fig. 2. SCAPS-1D simulation procedure.

Table 2
Input parameters of distinct HTLs.

Parameter	CuI [18,19]	CuSCN [19,20]	Cu_2O [15,19,21]	PEDOT: PSS [19,22,23]	CdTe [19,24]	$MASnBr_3$ [6,19,25]	CFTS [19,26]
Thickness (nm)	50	50	50	50	50	50	50
Eg (eV)	3.1	3.4	2.17	2.2	1.5	2.15	1.3
χ (eV)	2.1	1.9	3.2	2.9	3.9	3.39	3.3
ε	6.5	10	7.11	3.0	9.4	8.2	9.0
N_C (cm^{-3})	2.8×10^{19}	2.2×10^{19}	2.02×10^{17}	2.2×10^{15}	8.0×10^{17}	1×10^{18}	2.2×10^{18}
N_V (cm^{-3})	1×10^{19}	1.8×10^{18}	1.1×10^{19}	1.8×10^{18}	1.8×10^{19}	1×10^{18}	1.8×10^{19}
μ_n (cm^2/Vs)	100	1×10^{-4}	200	1×10^{-2}	3.2×10^{2}	1.6	21.98
μ_p (cm^2/Vs)	80	1×10^{-2}	80	2×10^{-4}	40	1.6	21.98
N_D (cm^{-3})	0	0	0	0	0	0	0
N_A (cm^{-3})	1×10^{18}	1×10^{16}	1×10^{18}	1×10^{19}	2×10^{14}	1×10^{18}	1×10^{18}
N_t (cm^{-3})	1×10^{14}	$1 \times 10^{14*}$	1×10^{14}	1×10^{14}	$1 \times 10^{14*}$	$1 \times 10^{14*}$	$1 \times 10^{14*}$

into training and testing sets for model training. To find the best hyperparameters, two distinct approaches were taken. First a parameter grid (a Python dictionary) inclusive various values for each parameter was prepared. The RandomizedSearchCV and GridSearchCV tools from Scikit-Learn were employed to assess the optimal combination of hyperparameters from this grid. The RandomizedSearchCV randomly selected combinations from the grid and evaluated their performance using cross-validation, while the GridSearchCV evaluated every possible combination of the provided parameters. To ensure the model's performance and consistency, five-fold cross-validation was performed on the final model. Subsequently, a separate test set was used to assess the model's accuracy and overall performance. Random Forest's ability to aggregate the results of individual decision trees helped mitigate overfitting issues and improved the model's predictive performance [30]. Overfitting occurs when a model becomes too tailored to the training data, resulting in high variance and poor predictive performance on the test set. To mitigate overfitting, various techniques can be applied, such as comparing the prediction results between the training and test sets. Several approaches can be used, including increasing the number of decision trees and reducing their maximum depth. The prediction's coefficient of determination acted as the evaluation metric. According to the Sklearn, this metric exhibited as is defined as:

$R^2 = \left(1 - \frac{u}{v}\right)$ (4).u represent the residual sum of squares $((y_{pred} - y_{true})^2)$. Meanwhile, v stands for the total sum of squares $= ((y_{pred} - \text{mean}(y_{true}))^2$. It's worth noting that a constant model that solely predicts the anticipated y, irrespective of input, could obtain a score of R^2 equal to 0.0. The significance of a node j within a single decision tree, the calculation involved as follows [30]:

$$ni_j = w_j C_j - w_{left(j)} Cleft(j) - w_{right(j)} C_{right(j)} \quad (5)$$

where w_j denotes the weighted number of samples in node j as a fraction of the total weighted number of samples. C_j represents the impurity in node j and left (j) and right (j) corresponds to the child nodes associated with it. The following calculation is used to determine feature's importance [30]:

$$fi_i = \frac{\sum_{j:nodejsplitsonfeaturei} ni_j}{\sum_{j \in allnodes} ni_j} \quad (6)$$

Ultimately, the feature values were calculated as the mean of every particular decision tree.

3. Results and discussion

3.1. HTL

The HTL exhibits an important role in optimization of PSC performance via two ways such as it transports the light generated holes towards the back contact and HTL hampers the interaction between absorber layer and back contact. Despite the comprehensive application of Spiro-OMeTAD into PSC to increase PCE, low hole mobility (μ_p), and environment instability confined its commercial application. Consequently, various kinds of HTLs like CuI, CuSCN, Cu_2O, PEDOT: PSS, CdTe, $MASnBr_3$ and CFTS have been analyzed shown in Table 2. To begin the simulation, a PSC architecture of FTO/TiO_2/$KSnI_3$/Spiro-OMeTAD/Au was simulated to validate with previously reported results [8]. WS_2 opted as ETL due to its high electron mobility (μ_n), wide range of E_g, and high conductivity of $10^{-3}\ \Omega^{-1}\ cm^{-1}$ [31]. The validation of the simulated result shown below in Table 3. The highest occupied molecular orbital (HOMO) of HTL should be appropriately aligned with that of the perovskite layer. The proper optimization is required to select appropriate ETL and HTL to maximize the performance of the PSC. The energy bandgap diagram of ETL, absorber layer and various HTLs illustrated below in Fig. 3. The J-V and QE plots of PSC with various HTLs are shown below in Fig. 4. The plotted curve from simulations is well-fitted with the previously reported outcomes, which validates numerical simulations.

Table 3
Comparison with previously reported results.

Cell structure	V_{OC} (V)	J_{SC} (mA/cm^2)	FF (%)	PCE (%)
FTO/TiO_2/$KSnI_3$/Spiro-OMeTAD/W	1.70	15.85	36.13	9.77 [8]
FTO/C_{60}/$KSnI_3$/PTAA/C	0.76	17.44	80.80	10.83 [32]
FTO/TiO_2/$KSnI_3$/ Spiro-OMeTAD/Au	1.48	13.20	79.07	15.55 [This work]

To examine the impact of the several HTLs like Spiro-OMeTAD, CuI, CuSCN, Cu_2O, PEDOT: PSS, CdTe, $MASnBr_3$ and CFTS on the PV parameters of PSC. The thickness of HTLs was 50 nm throughout the simulation. The simulated results are demonstrated in Fig. 4. The PV parameters with several HTLs are presented below in Table 4.

The Spiro-OMeTAD and PEDOT: PSS both are organic materials. The E_g for Spiro-OMeTAD is 3 eV and for PEDOT: PSS is 2.2 eV depicted in table 1 and 2. The PEDOT: PSS has excellent properties such as high electrical conductivity, mechanical flexibility and suitable for thin film formation [33]. The simulation results demonstrated that $KSnI_3$-based PSC exhibited maximum PCE of 21.08 % (with V_{OC} of 1.59 V, J_{SC} of 15.08 mA/cm^2 and FF of 87.43 %) for WS_2 as ETL and PEDOT: PSS employed as HTL. The HOMO level of Spiro-OMeTAD and PEDOT: PSS has good band alignment with $KSnI_3$ therefore with PEDOT: PSS it has achieved maximum PCE but not for Spiro-OMeTAD. This can be attributed to lower hole mobility (μ_p) than PEDOT: PSS [34,35]. Hence, PEDOT: PSS is an appropriate HTL for achieving high PCE. Furthermore, the PEDOT: PSS exhibits good stability in air over Spiro-OMeTAD so that it has wide application in PSCs. Zuo et al. (2017) doped PEDOT: PSS with polymer electrolyte PSS-Na which results in high work function and good band alignment of modified PEDOT: PSS with the absorber layer. Hence, enhanced PCE has been obtained from 12.35 % to 15.56 % [36]. The small amount of 2,3,5,6-tetrafluoro-7,7,8,8-tetracyanoquinodimethane (F4-TCNQ) added to PEDOT: PSS could improve the electrical conductivity and better band matching with the absorber layer. Therefore, the PCE enhanced from 13.30 % to 17.22 % [37,38].

3.2. Optimization of HTL

3.2.1. Impact of thickness

In the preceding section, Table 4 highlighted that PEDOT: PSS emerges as the most promising HTL for $KSnI_3$-based PSC. As the attributes of HTL, such as thickness, doping density (N_A), and defect density (N_t), exert a substantial influence on PSC performance and these parameters must be optimized. Notably, HTL thickness exhibits a vital role in facilitating efficacious carrier transport and controlling the recombination rate [39]. Consequently, the thickness of an HTL is optimized to get high PCE. Fig. 5a, b demonstrates the impact of PEDOT: PSS thickness on the J_{SC} and V_{OC} of PSCs. The findings reveal that increasing PEDOT: PSS thickness up to 250 nm enhances both J_{SC} and V_{OC}, beyond which they become constant. According to experimental surveys, thicker HTLs yields smoother film surfaces, reducing interfacial recombination and thus enhancing V_{OC}. The suitable HTL thickness is required to avoid series resistance (R_S) for thicker layers [40,41]. Firstly, a thinner HTL might not accumulate a sufficient number of holes. Secondly, the region among the absorber layer and a thin HTL might not be as potent as in thicker layers, leading to inadequate driving force for carrier partition. The study also indicates that when HTL thickness exceeds 250 nm both V_{OC} and J_{SC} tend to be constant. This is primarily due to carrier travel distance to reach the interface. Consequently, traps promote non-radiative Shockley–Read–Hall (SRH) recombination and enhanced trap-assisted recombination rate [42]. The influence of thickness on PV parameters is depicted in Fig. 5. The initial fall in the

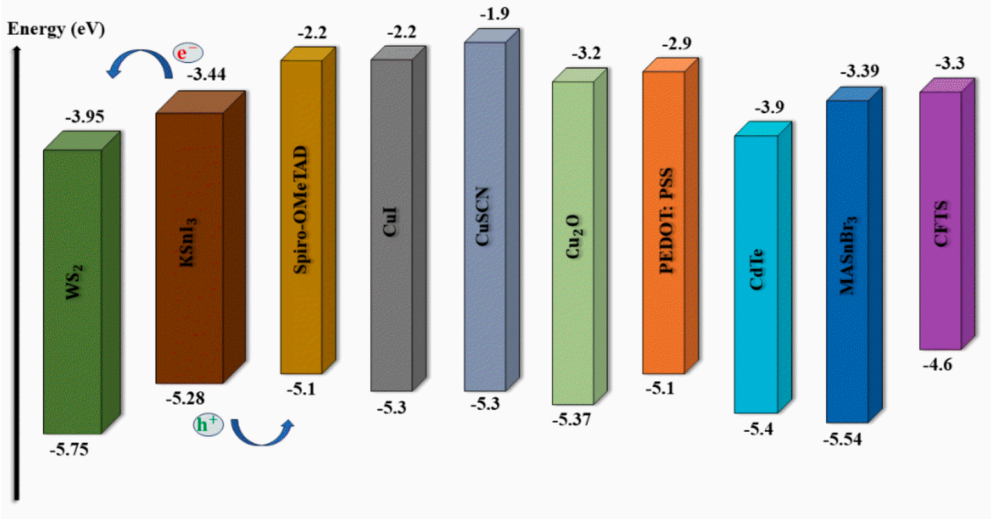


Fig. 3. Energy band diagram of various HTLs.

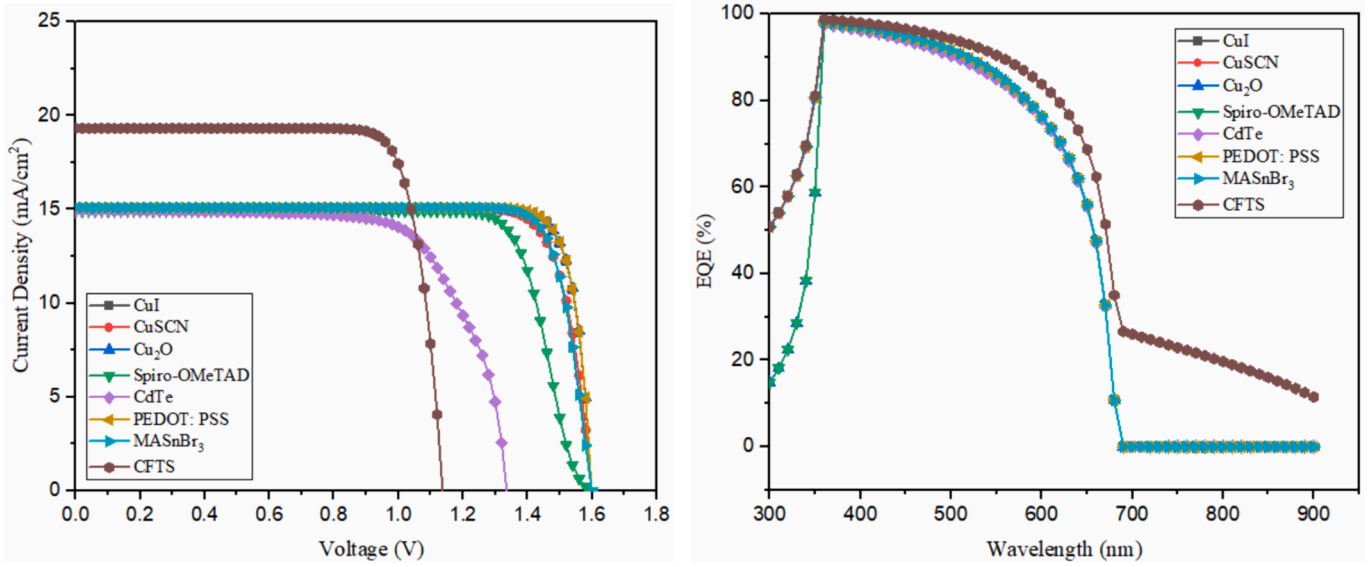


Fig. 4. (**a**) J-V and (**b**) QE plots of simulated PSCs with various HTLs.

Table 4
The PV parameters of $KSnI_3$-based PSC with distinct HTLs.

Cell structure	V_{OC} (V)	J_{SC} (mA/cm^2)	FF (%)	PCE (%)
FTO/WS_2/$KSnI_3$/CuI/Au	1.59	15.01	87.40	20.98
FTO/WS_2/$KSnI_3$/CuSCN/Au	1.59	15.01	84.48	20.27
FTO/WS_2/$KSnI_3$/Cu_2O/Au	1.59	14.96	87.63	20.96
FTO/WS_2/$KSnI_3$/Spiro-OMeTAD/Au	1.59	14.89	79.18	18.85
FTO/WS_2/$KSnI_3$/PEDOT: PSS/Au	1.59	15.08	87.43	21.08
FTO/WS_2/$KSnI_3$/CdTe/Au	1.33	14.93	71.06	14.18
FTO/WS_2/$KSnI_3$/$MASnBr_3$/Au	1.59	15.09	85.14	20.56
FTO/WS_2/$KSnI_3$/CFTS/Au	1.13	19.32	81.26	17.85

curve is a result of an increase in R_S. The optimized thickness of HTL is 250 nm, yielding a PCE of 21.30 % demonstrated in Fig. 5d.

3.2.2. *Impact of N_A*

The findings from previous studies have shown the significant impact of doping density (N_A) in the HTL vastly affects the PV parameters of PSCs, including conductivity, recombination rate and V_{OC} [6]. Hence, the solar cell parameters were simulated in the range of N_A values from 10^{14} to 10^{22} cm^{-3} at a thickness of 250 nm of HTL. As depicted in Fig. 6a, an enhancement in N_A leads to an increase in V_{OC}. In layered semiconductor architectures, the alignment of the fermi level at the same value results in an equilibrium state, causing band bending and built-in voltage (V_{bi}) [43]. This internal electric field drifts segregated photo-generated e$^-$s and h$^+$s towards the n and p regions respectively, where they gather. Under illumination and open-circuit conditions, the excited carriers induce the separation of quasi-Fermi-levels (EFn and EFp) for e$^-$s and h$^+$s, respectively, within the photoactive semiconductor material. This effectively nullifies V_{bi}, and forms a photo-voltage referred to as V_{OC} [44].

Raising the dopant concentration leads to reduction in the recombination rate, owing to an enhanced internal electric field and accelerated carrier segregation. Simultaneously, there's a slight decrease in J_{SC} within the N_A range of 10^{14} to 10^{18} cm^{-3}. Notably, a pronounced fall in the J_{SC} is observed at 10^{19} cm^{-3} due to generation of numerous deep Coulomb traps that reduce the μ_p [45]. The Fig. 6a shows that the maximum PCE of 21.32 % is obtained at N_A of 10^{20} cm^{-3}.

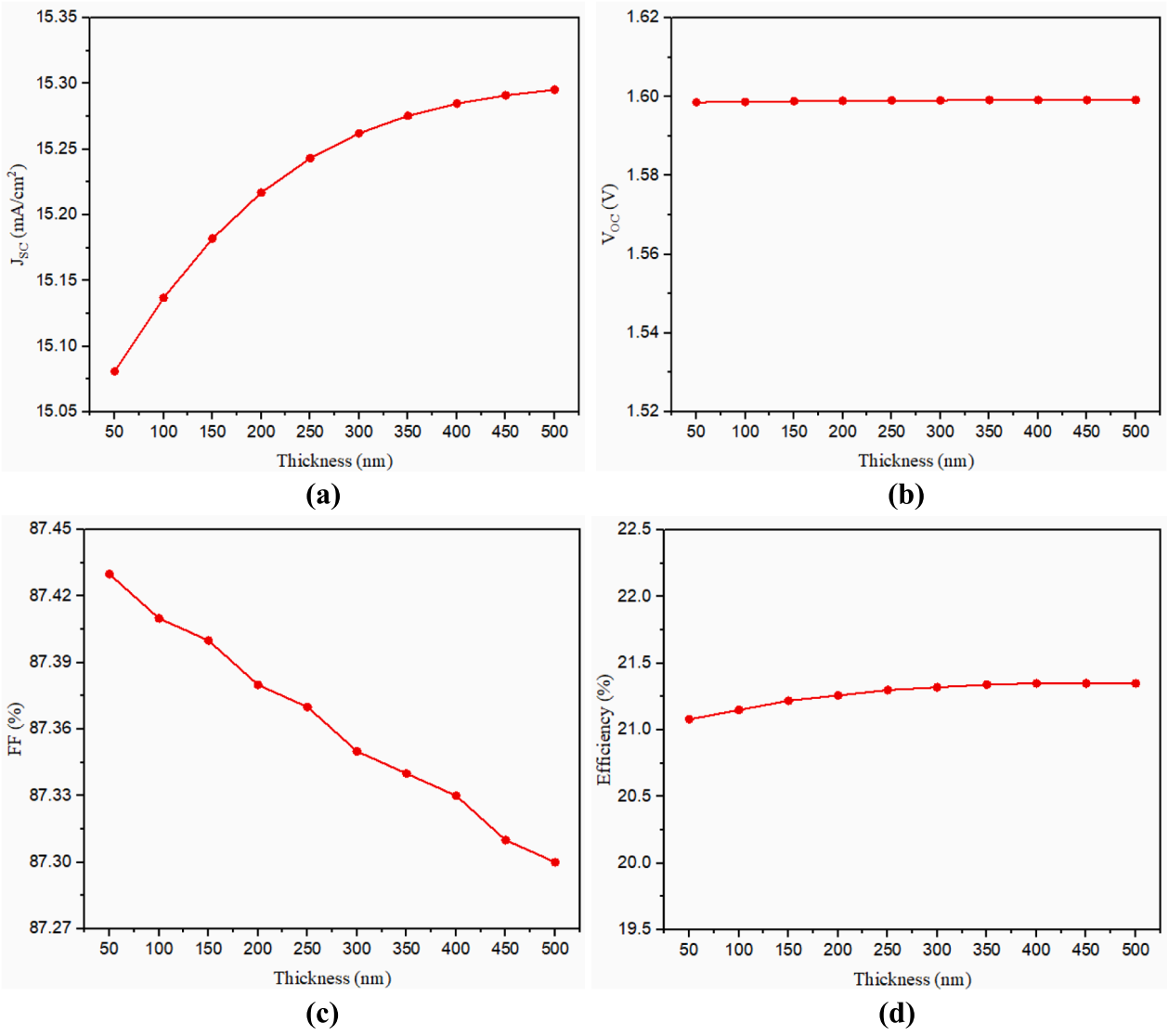


Fig. 5. Impact of the thickness of PEDOT: PSS layer on PV parameters of PSC.

3.2.3. *Impact of N_t*

In addition to the N_A, HTL defect density (Nt) affects the performance of PSCs. The impact on the PSC characteristics demonstrated in Fig. 7. The N_t of 10^{17} to 10^{22} cm^{-3}, Fig. 7a, b, c and d reveal no significant changes in PV parameters. However, a pronounced decline in V_{OC}, J_{SC} and PCE surface for change in N_t from 10^{14} cm^{-3} to 10^{17} cm^{-3}. This precipitous shift is primarily attributed to the rapid emergence of numerous recombination sites in the HTL and at interfaces [46]. A high value of N_t in HTL, arising from diverse sources like undesirable foreign atoms, and native defects, give rise to shallow or deep traps. These traps adversely impact cell performance. Fig. 7d shows that an increase in interface defects leads to reduced efficiency. Beyond a certain N_t, it becomes constant. These interface defects, primarily stemming from the lattice mismatch between absorber layer (AL) and HTL, yield deep traps at the interface. These traps function as SRH recombination centers. Consequently, PCE experiences a decline, reaching about 20.97 % at Nt $= 10^{22}$ cm^{-3}. In the beginning FF decline is attributed to the abundance of recombination centers and R_S resulting from a significant number of traps (Fig. 7c). Despite the superior outcomes at lower N_t, simulation results diverge from experimental findings. Hence, a value of 1×10^{14} cm^{-3} was selected.

3.3. *Optimization of $KSnI_3$ layer*

The above outcomes have shown that PEDOT: PSS is the most suitable HTL for $KSnI_3$-based PSC. The maximum PCE of beyond 21 % has been obtained at 250 nm of thickness of HTL, N_A of 10^{20} cm^{-3} and at N_t of 1×10^{14} cm^{-3}. Therefore, we have opted for this optimized value of thickness, N_A and N_t for the further simulation study to optimize the performance of $KSnI_3$-based PSC.

3.3.1. *Impact of thickness*

The characteristics parameters like J_{SC}, V_{OC}, FF and PCE of PSCs are significantly influenced by the thickness of the perovskite absorber layer. In our investigation, we explored the impact of thickness of the absorber layer on the performance of the architecture: FTO/WS_2(100 nm)/$KSnI_3$/PEDOT: PSS (250 nm)/Au. The thickness of the $KSnI_3$ layer has exhibited a straight relation with the evaluation of the outcomes of PSC through its effect on the charge carrier's diffusion length. It is known to us that as the thickness of the layer decreases, the absorption rate also lowers. This results in a decrement in photocurrent and therefore PCE of the PSC also reduces from 200 to 1000 nm. The change in PCE is depicted below in Fig. 8a. However, with the thickness increasing up to 550 nm, PCE gradually rises. This phenomenon can be attributed to the generation of large amounts of charge carriers by

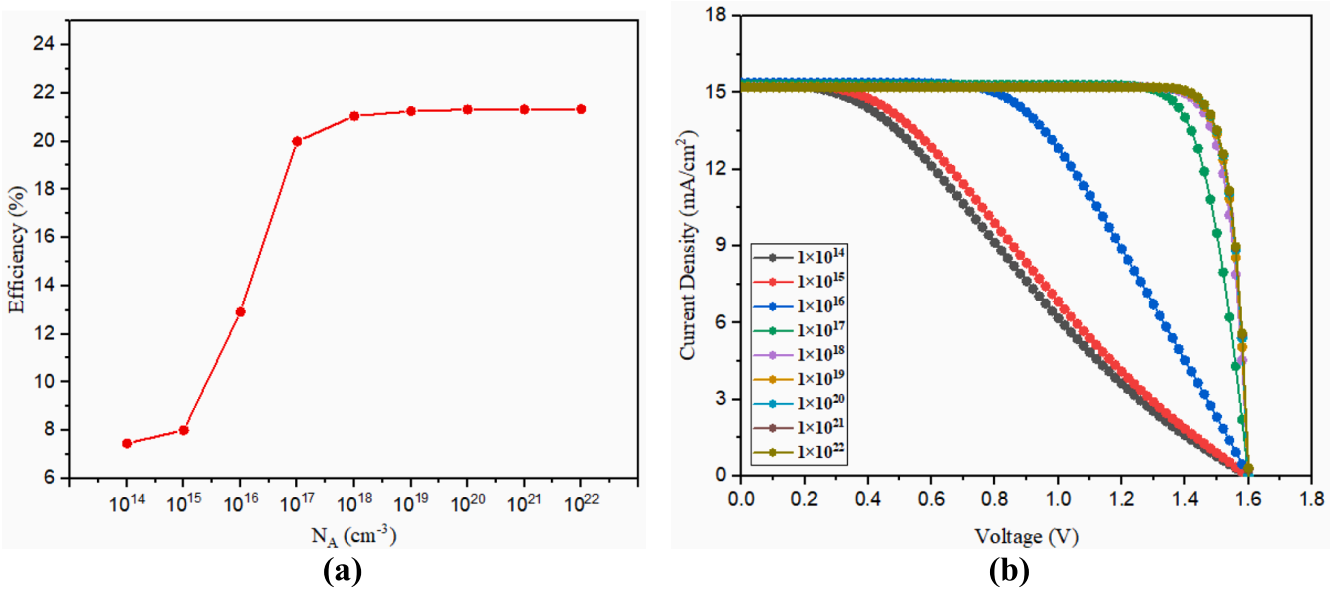


Fig. 6. (**a**) Efficiency versus N_A of the PEDOT: PSS. (**b**) J-V curves of HTL depending on its N_A. (**c**) Efficiency vs thickness of the HTL and (**d**) Effect of N_t on PCE.

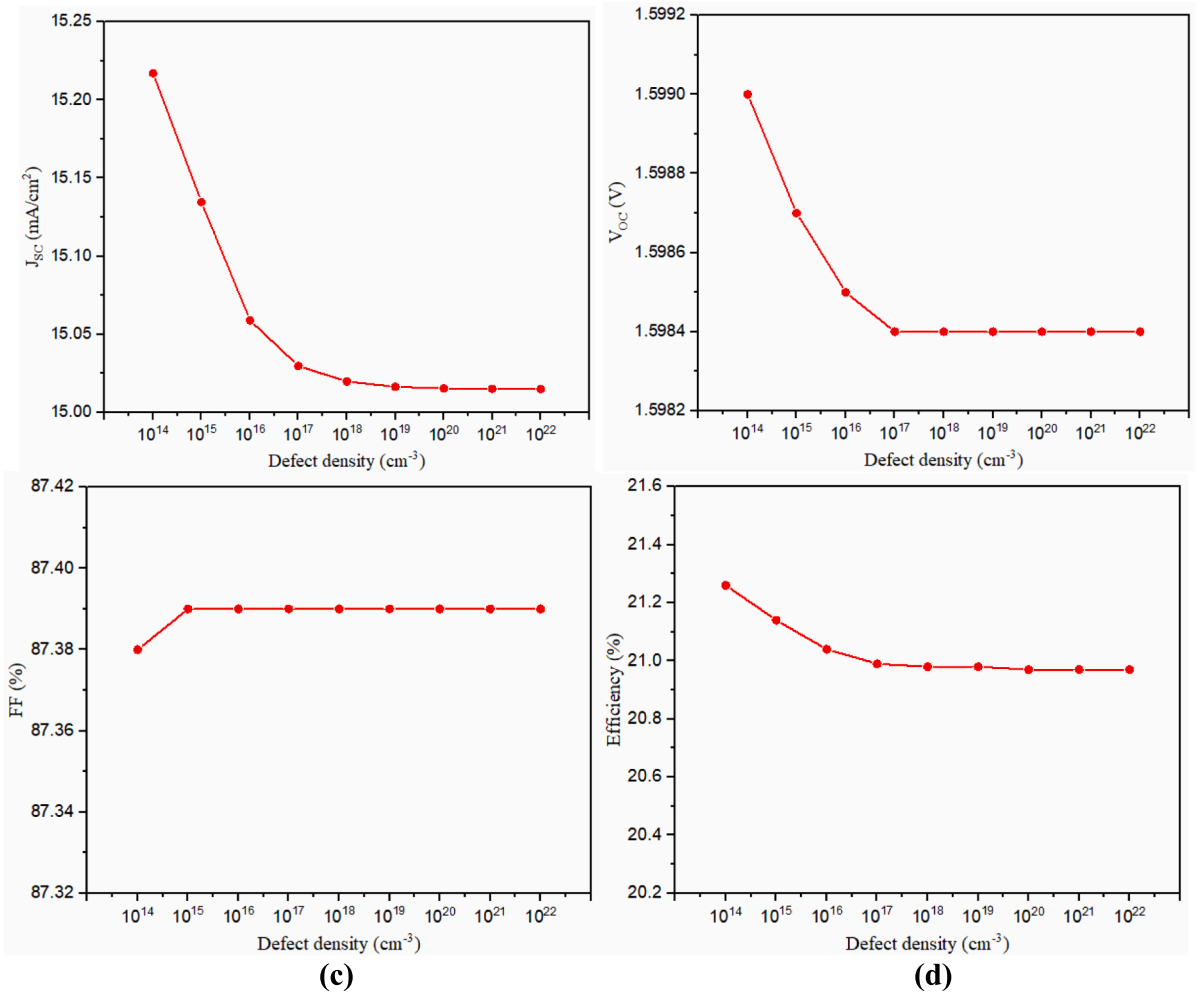


Fig. 7. Impact of N_t of the PEDOT: PSS on PV parameters of $KSnI_3$-based PSC.

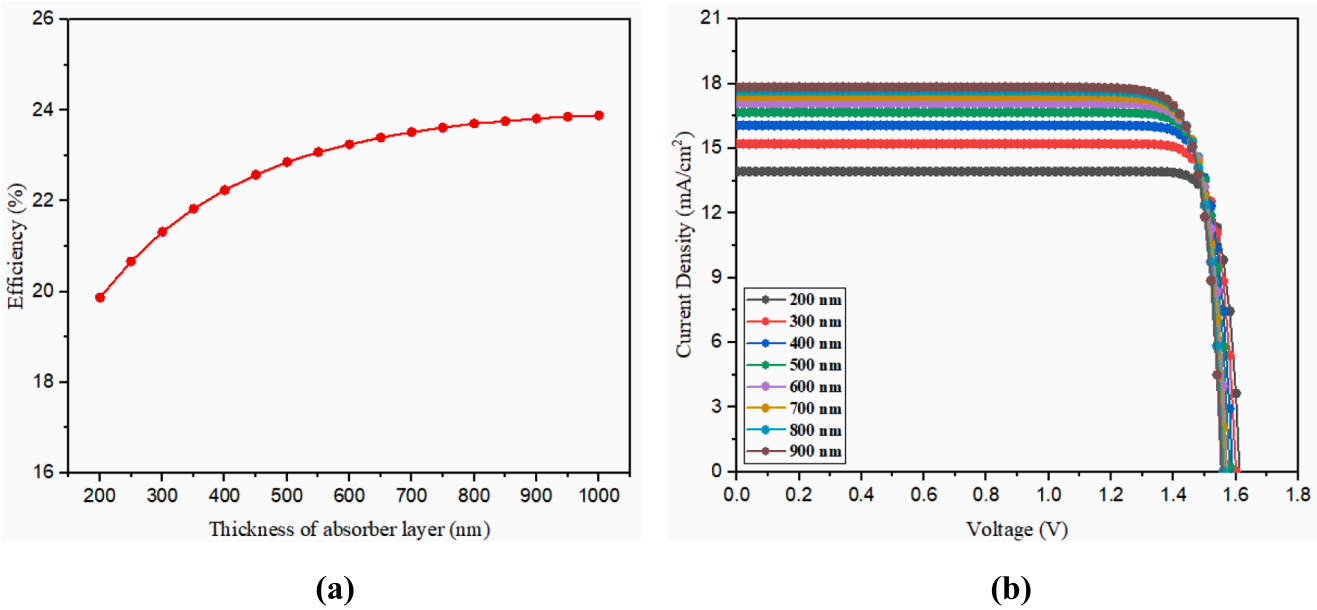


Fig. 8. (**a**) Efficiency vs thickness of the absorber layer. (**b**) J-V curves with function of absorber layer thickness.

absorbed photons, leading to higher EQE, J_{SC}, and overall improved performance [47]. This enhancement is also evident in the J-V curves (Fig. 8b). As the absorber layer thickness exceeds 550 nm, both V_{OC} and FF parameters decrease. Optimum thickness value for $KSnI_3$-based PSC is 550 nm. The maximum efficiency goes up to 23.08 % together with V_{OC} is 1.57 V, J_{SC} is 16.90 mA/cm^2 and FF is 86.59 % at a thickness of 550 nm.

3.3.2. *Impact of N_t*

Similarly, with the HTL influence, there is a lack of substantial change in the PV parameters until an N_t value of 10^{13} cm^{-3} is reached. These PV parameters exhibit a decline when a high defect concentration is present. A heightened defect concentration within a low-quality perovskite layer results in an increased number of non-radiative recombination centers, consequently deteriorating PV parameters of PSC. This degradation can be attributed to alterations in the diffusion length of charge carriers and their lifetime. While a low N_t value might contribute to high PCE in simulated PSCs, the elevated fabrication cost is a deterrent. Moreover, a low N_t value does not accurately reflect real-world conditions due to the instability of such perovskites under humid conditions, leading to partial degradation. Therefore, we propose N_t values ranging from 10^{13} to 10^{15} cm^{-3} for the building of high-performance PSCs with reasonable stability. The optimized N_t is 1×10^{13} cm^{-3} for $KSnI_3$ perovskite layer at which it has achieved the maximum PCE of 23.08 %, as demonstrated in Fig. 9a. The J-V plot of $KSnI_3$ PSCs based on the N_t of the absorber layer is simulated and demonstrated in Fig. 9b. For perovskite layers when N_t is increased, the undesirable recombination rate also rises, attributed to the formation of defects that act as dangling bonds, serving as trap state for the charge carriers. This enhancement in recombination leads to fall in J_{SC}, which is subsequently responsible for the decrement in FF of solar devices. The device PCE also decreased from 23.08 % to 7.85 %, when N_t was increased from 10^{13} to 10^{20} cm^{-3}. Such an increasing amount of recombination resulted in a sharp fall in PCE of PSC.

3.4. *Optimized device*

The optimized ETL and HTL are WS_2 and PEDOT: PSS respectively with the Au (gold) as metal back contact. The optimized parameters such as 1) Thickness: PEDOT: PSS (250 nm), $KSnI_3$ (550 nm), WS_2 (100 nm) and FTO (400 nm); 2) N_A: PEDOT: PSS (10^{20} cm^{-3}) and 3) N_t: PEDOT: PSS (1×10^{14} cm^{-3}), $KSnI_3$ (1×10^{13} cm^{-3}) and WS_2 (1×10^{13} cm^{-3}). The PSC structure and energy band diagram of the optimized device demonstrated in Fig. 10. The optimized PCE of 23.08 % with V_{OC} = 1.57 V, J_{SC} = 16.90 mA/cm^2, and FF of 86.59 % has been obtained. The JV and QE curves of optimized PSC are illustrated in Fig. 11(a-b).

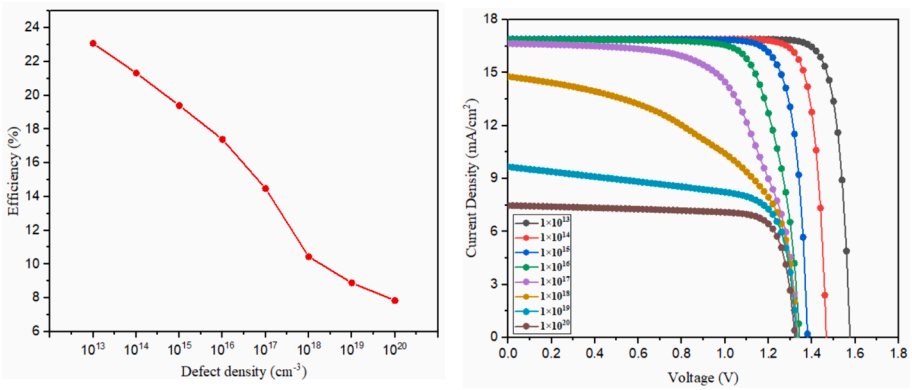


Fig. 9. (**a**) Efficiency vs N_t of the absorber layer and (**b**) Effects of N_t on the J-V characteristics of $KSnI_3$-based PSCs.

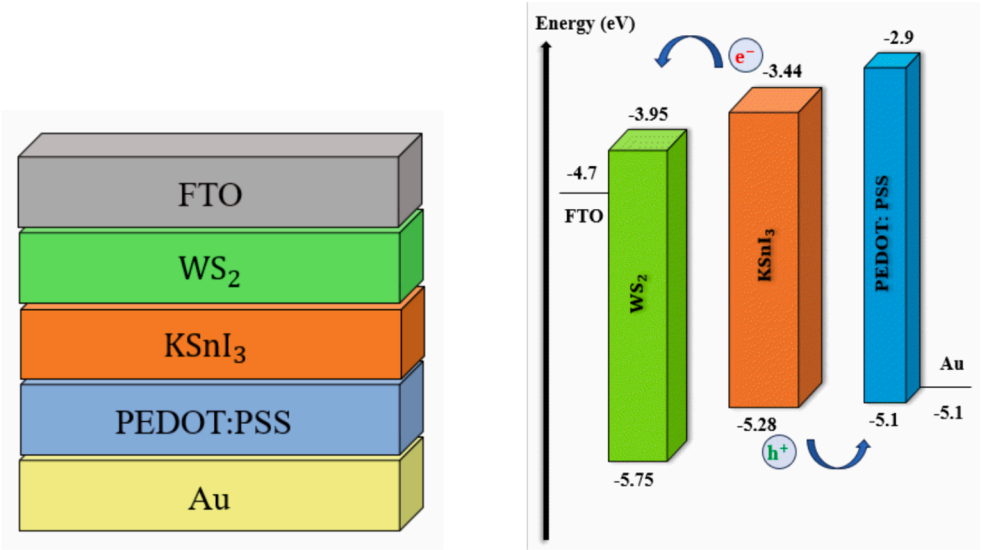


Fig. 10. Architecture and band alignment diagram of optimized PSC.

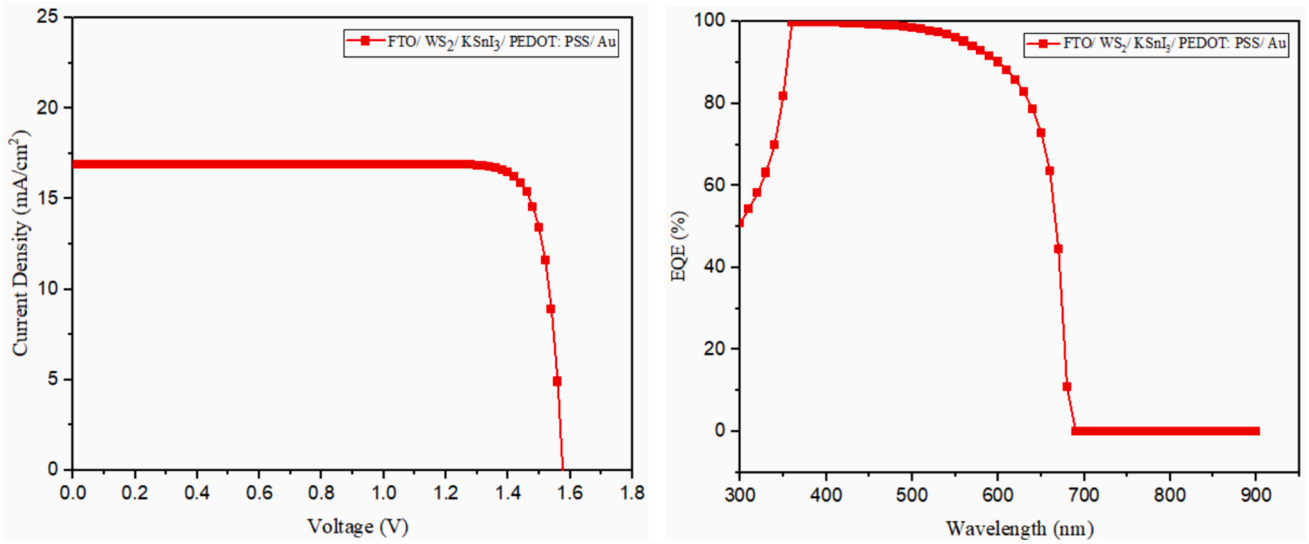


Fig. 11. (**a**) J-V and (**b**) QE plot of optimized device.

3.5. *Machine learning*

In this work, python is employed for the prediction of factors affecting the performance of the PSC. Due to its robustness to overfitting and optimal complexity, the Random Forest algorithm was used for the training and testing [48]. The researchers noted that more powerful rivals, like neural networks and deep neural networks, acquire needless complexity, leading to overfitting and reduced performance [49,50]. Considering the relatively small size of the dataset and its high complexity, compiling a larger set of records would be beneficial. The research focused on reducing the dataset's complexity to improve the model's accuracy. By employing Scikit-Learn, influence of every feature on the entire predictive performance of the ML model was observed. The eleven most important features were identified, and their relative importance was shown in Fig. 12. It has shown that the features such as thickness, N_t, N_A, of the HTL were found to be the most responsible in differentiating between the performance of two distinct PSCs. The outcomes guide for the design of new solar cells. Subsequently, a new model was built using the selected important features aiming to achieve a similar level of performance while reducing the model's size due to the

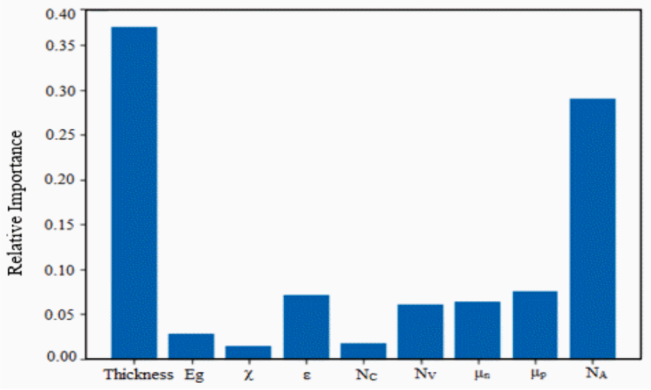


Fig. 12. Different features with relative importance in the ML model.

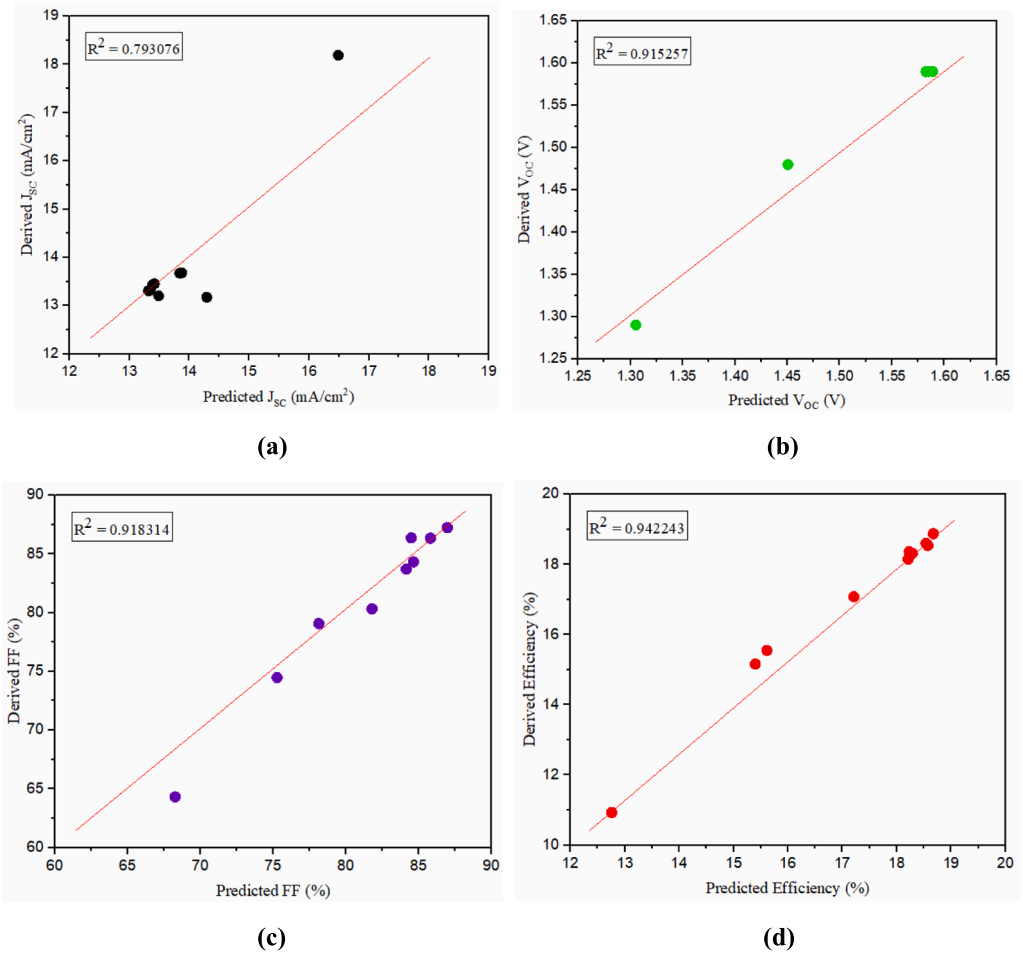


Fig. 13. Derived and predicted values of PV parameters.

fewer features used. The model results were measured and illustrated in Fig. 13 (a-d). Ideally, the slope of plots is near to 1, indicating accurate predictions. The ML model achieved a mean accuracy of 0.89 % and mean R^2 of 0.8921 between actual and predicted values for each target feature. While the accuracy might not be extremely high, the results are encouraging for further investigation and bridging the gap between artificial intelligence and materials science. The model trained for large input data which leads to highly efficient and realistic results.

4. Conclusions

This work comprises analysis of $KSnI_3$-based PSC with the various HTLs. The effect of distinct parameters, including thickness of perovskite absorber layer and HTL, as well as N_A and N_t have been examined using SCAPS-1D. The ML model was used to predict the performance of the PSCs. The study demonstrated PEDOT: PSS as a promising HTL that could replace the expensive and less-conductive Spiro-OMeTAD. The maximum PCE of 23.08 % was obtained for the configuration of FTO/WS_2/$KSnI_3$/PEDOT: PSS/Au. The SCAPS-1D outcomes, deduced that the optimized thickness of HTL was 200 nm, with N_t of 1×10^{14} cm^{-3} and a N_A of 1×10^{20} cm^{-3} results in maximum PCE. The optimized thickness of the $KSnI_3$ layer is 550 nm and at N_t of 1×10^{13} cm^{-3} for an efficient PSC. The incorporation of the ML approach provided enhanced insight into the critical factors involved in designing and fabricating solar cells. The ML model demonstrated an accuracy score of 89 % in predicting the performance metrics of PSCs. The device PCE improved beyond 23 % by optimizing parameters such as thickness of absorber layer and HTL, N_t and N_A. The study focused on proper optimization of perovskite absorber layer and HTL for advanced and efficient architecture of $KSnI_3$-based PSC. A precise and powerful model can be trained by providing a large number of inputs to ML. With the series of this study, we can study the lead free and other potential material to optimize the performance of PSCs using the ML model.

CRediT authorship contribution statement

Rahul Kundara: Conceptualization, Data curation, Investigation, Methodology, Software, Validation, Writing – original draft. **Sarita Baghel:** Formal analysis, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The author, Rahul Kundara is thankful to Delhi Technological University, New Delhi for providing a junior research fellowship. We are very grateful to Marc Burgelman, the Electronics and Information

Systems (ELIS), University of Gent, Belgium, for providing us with SCAPS-1D software.

References

[1] A. Singha, A. Paul, S. Koul, V. Sharma, S. Mallick, K.R. Balasubramaniam, D. Kabra, Stable and efficient large area 4T Si/perovskite tandem photovoltaics with sputtered transparent contact, Solar RRL. 7 (2023) 2300117, https://doi.org/10.1002/solr.202300117.

[2] N. Rai, S. Rai, P.K. Singh, P. Lohia, D.K. Dwivedi, Analysis of various ETL materials for an efficient perovskite solar cell by numerical simulation, J. Mater. Sci. Mater. Electron. 31 (2020) 16269–16280, https://doi.org/10.1007/s10854-020-04175-z.

[3] I. Elango, M. Selvamani, PC. Ramamurthy, AV. Kesavan, Studying VOC in lead free inorganic perovskite photovoltaics by tuning energy bandgap and defect density, Ceram. Int. 48 (2022) 29414-29420. https://doi. org/10.1016/j.ceramint.2022.06.125.

[4] M.S. Islam, K. Sobayel, A. Al-Kahtani, M.A. Islam, G. Muhammad, N. Amin, M. Shahiduzzaman, M. Akhtaruzzaman, Defect study and modelling of SnX3-based perovskite solar cells with SCAPS-1D, Nanomaterials 11 (2021), https://doi.org/10.3390/nano11051218.

[5] N. Lakhdar, A. Hima, Electron transport material effect on performance of perovskite solar cells based on CH3NH3GeI3, Opt Mater (AMST). 99 (2020), https://doi.org/10.1016/j.optmat.2019.109517.

[6] S. Mushtaq, S. Tahir, A. Ashfaq, R. Sebastian Bonilla, M. Haneef, R. Saeed, W. Ahmad, N. Amin, Performance optimization of lead-free MASnBr 3 based perovskite solar cells by SCAPS-1D device simulation, Sol. Energy 249 (2023) 401–413, https://doi.org/10.1016/j.solener.2022.11.050.

[7] N.K. Noel, S.D. Stranks, A. Abate, C. Wehrenfennig, S. Guarnera, A.A. Haghighirad, A. Sadhanala, G.E. Eperon, S.K. Pathak, M.B. Johnston, A. Petrozza, L.M. Herz, H. J. Snaith, Lead-free organic-inorganic tin halide perovskites for photovoltaic applications, Energy Environ. Sci. 7 (2014) 3061–3068, https://doi.org/10.1039/c4ee01076k.

[8] M.G. Ju, M. Chen, Y. Zhou, H.F. Garces, J. Dai, L. Ma, N.P. Padture, X.C. Zeng, Earth-abundant nontoxic titanium(IV)-based vacancy-ordered double perovskite halides with Tunable 1.0 to 1.8 eV bandgaps for photovoltaic applications, ACS Energy Lett. 3 (2018) 297–304, https://doi.org/10.1021/acsenergylett.7b01167.

[9] Z. Zhang, X. Li, X. Xia, Z. Wang, Z. Huang, B. Lei, Y. Gao, High-quality (CH3NH3)3Bi2I9 film-based solar cells: pushing efficiency up to 1.64%, J. Phys. Chem. Lett. 8 (2017) 4300–4307, https://doi.org/10.1021/acs.jpclett.7b01952.

[10] M. Wang, P. Zeng, S. Bai, J. Gu, F. Li, Z. Yang, M. Liu, High-quality sequential-vapor-deposited Cs2AgBiBr 6 thin films for lead-free perovskite solar cells, Solar RRL. 2 (2018), https://doi.org/10.1002/solr.201800217.

[11] Y. Zong, Y. Zhou, Y. Zhang, Z. Li, L. Zhang, M.G. Ju, M. Chen, S. Pang, X.C. Zeng, N. P. Padture, Continuous grain-boundary functionalization for high-efficiency perovskite solar cells with exceptional stability, Chem 4 (2018) 1404–1415, https://doi.org/10.1016/j.chempr.2018.03.005.

[12] G. Pindolia, S.M. Shinde, P.K. Jha, Non-leaded, KSnI3 based perovskite solar cell: A DFT study along with SCAPS simulation, Mater. Chem. Phys. 297 (2023), https://doi.org/10.1016/j.matchemphys.2023.127426.

[13] A. Tara, V. Bharti, S. Sharma, R. Gupta, Device simulation of FASnI3 based perovskite solar cell with Zn(O0.3, S0.7) as electron transport layer using SCAPS-1D, Opt Mater (amst) 119 (2021), https://doi.org/10.1016/j.optmat.2021.111362.

[14] W. Yan, Y. Liu, Y. Zang, J. Cheng, Y. Wang, L. Chu, X. Tan, L. Liu, P. Zhou, W. Li, Z. Zhong, Machine learning enabled development of unexplored perovskite solar cells with high efficiency, Nano Energy 99 (2022) 107394, https://doi.org/10.1016/j.nanoen.2022.107394.

[15] S. Rai, B.K. Pandey, D.K. Dwivedi, Modeling of highly efficient and low cost CH3NH3Pb(I1-xClx)3 based perovskite solar cell by numerical simulation, Opt Mater (amst). 100 (2020), https://doi.org/10.1016/j.optmat.2019.109631.

[16] M. Kumar, A. Raj, A. Kumar, A. Anshul, An optimized lead-free formamidinium Sn-based perovskite solar cell design for high power conversion efficiency by SCAPS simulation, Opt Mater (amst). 108 (2020), https://doi.org/10.1016/j.optmat.2020.110213.

[17] M.K. Hossain, G.F.I. Toki, A. Kuddus, M.H.K. Rubel, M.M. Hossain, H. Bencherif, M.F. Rahman, M.R. Islam, M. Mushtaq, An extensive study on multiple ETL and HTL layers to design and simulation of high-performance lead-free CsSnCl3-based perovskite solar cells, Sci. Rep. 13 (2023), https://doi.org/10.1038/s41598-023-28506-2.

[18] A.A. Kanoun, M.B. Kanoun, A.E. Merad, S. Goumri-Said, Toward development of high-performance perovskite solar cells based on CH3NH3GeI3 using computational approach, Sol. Energy 182 (2019) 237–244, https://doi.org/10.1016/j.solener.2019.02.041.

[19] S. Bhattarai, T.D. Das, Optimization of carrier transport materials for the performance enhancement of the MAGeI3 based perovskite solar cell, Sol. Energy 217 (2021) 200–207, https://doi.org/10.1016/j.solener.2021.02.002.

[20] S.Z. Haider, H. Anwar, M. Wang, Theoretical device engineering for high-performance perovskite solar cells using CuSCN as hole transport material boost the efficiency above 25%, Physica Status Solidi (A) Applications and Materials Science 216 (2019) https://doi.org/10.1002/pssa.201900102.

[21] B. Li, H. Di, B. Chang, R. Yin, L. Fu, Y.N. Zhang, L. Yin, Efficient passivation strategy on Sn related defects for high performance all-inorganic CsSnI3 perovskite solar cells, Adv. Funct. Mater. 31 (2021), https://doi.org/10.1002/adfm.202007447.

[22] U. Mandadapu, S.V. Vedanayakam, K. Thyagarajan, Simulation and analysis of lead based perovskite solar cell using SCAPS-1D, Indian, J. Sci. Technol. 10 (2017) 1–8, https://doi.org/10.17485/ijst/2017/v10i11/110721.

[23] K. Deepthi Jayan, V. Sebastian, Comprehensive device modelling and performance analysis of MASnI3 based perovskite solar cells with diverse ETM, HTM and back metal contacts, Sol. Energy 217 (2021) 40–48, https://doi.org/10.1016/j.solener.2021.01.058.

[24] O. Ahmad, A. Rashid, M.W. Ahmed, M.F. Nasir, I. Qasim, Performance evaluation of Au/p-CdTe/Cs2TiI6/n-TiO2/ITO solar cell using SCAPS-1D, Opt Mater (amst). 117 (2021), https://doi.org/10.1016/j.optmat.2021.111105.

[25] Y. Gan, X. Bi, Y. Liu, B. Qin, Q. Li, Q. Jiang, P. Mo, Numerical investigation energy conversion performance of tin-based perovskite solar cells using cell capacitance simulator, Energies (basel). 13 (2020), https://doi.org/10.3390/en13225907.

[26] M.K. Hossain, A.A. Arnab, R.C. Das, K.M. Hossain, M.H.K. Rubel, M.F. Rahman, H. Bencherif, M.E. Emetere, M.K.A. Mohammed, R. Pandey, Combined DFT, SCAPS-1D, and wxAMPS frameworks for design optimization of efficient Cs2BiAgI6-based perovskite solar cells with different charge transport layers, RSC Adv. 12 (2022) 34850–34873, https://doi.org/10.1039/d2ra06734j.

[27] S.P. Ong, W.D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K.A. Persson, G. Ceder, Python Materials genomics (pymatgen): a robust, open-source python library for materials analysis, Comput. Mater. Sci 68 (2013) 314–319, https://doi.org/10.1016/j.commatsci.2012.10.028.

[28] L. Ward, A. Dunn, A. Faghaninia, N.E.R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K.A. Persson, G. J. Snyder, I. Foster, A. Jain, Matminer: An open source toolkit for materials data mining, Comput. Mater. Sci 152 (2018) 60–69, https://doi.org/10.1016/j.commatsci.2018.05.018.

[29] R.T. Yarlagadda, Novateur publications international journal of innovations in engineering research and technology [IJIERT] python engineering automation to advance artificial intelligence and machine learning systems, 5 (2018). https://ssrn.com/abstract=3797347.

[30] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, B. Qiang, RFRSF: Employee turnover prediction based on random forests and survival analysis. In: Web Information Systems Engineering–WISE 2020: 21st International Conference, Amsterdam, The Netherlands, October 20–24, 2020, Proceedings, Part II 21 2020 (pp. 503-515). Springer International Publishing.

[31] K. Sobayel, M. Akhtaruzzaman, K.S. Rahman, M.T. Ferdaous, Z.A. Al-Mutairi, H. F. Alharbi, N.H. Alharthi, M.R. Karim, S. Hasmady, N. Amin, A comprehensive defect study of tungsten disulfide (WS2) as electron transport layer in perovskite solar cells by numerical simulation, Results Phys. 12 (2019) 1097–1103, https://doi.org/10.1016/j.rinp.2018.12.049.

[32] G. Pindolia, S.M. Shinde, Effect of organic charge transport layers on unleaded KSnI3 based perovskite solar cell, Results in Optics. 12 (2023) 100469, https://doi.org/10.1016/j.rio.2023.100469.

[33] Y. Xia, S. Dai, Review on applications of PEDOTs and PEDOT:PSS in perovskite solar cells, J. Mater. Sci. Mater. Electron. 32 (2021) 12746–12757, https://doi.org/10.1007/s10854-020-03473-w.

[34] L. Zhao, J. Mou, L. Zhu, J. Song, Synergistic effect of NiO and Spiro-OMeTAD for hole transfer in perovskite solar cells, J. Electron. Mater. 50 (2021) 6512–6517, https://doi.org/10.1007/s11664-021-09193-z.

[35] Q. Du, Z. Shen, C. Chen, F. Li, M. Jin, H. Li, C. Dong, J. Zheng, M. Ji, M. Wang, Spiro-OMeTAD:Sb2S3 hole transport layer with triple functions of overcoming lithium salt aggregation, long-term high conductivity, and defect passivation for perovskite solar cells, Solar RRL. 5 (2021), https://doi.org/10.1002/solr.202100622.

[36] C. Zuo, L. Ding, Modified PEDOT layer makes a 1.52 V Voc for perovskite/PCBM solar cells, Adv. Energy Mater. 7 (2017), https://doi.org/10.1002/aenm.201601193.

[37] D. Liu, Y. Li, J. Yuan, Q. Hong, G. Shi, D. Yuan, J. Wei, C. Huang, J. Tang, M. K. Fung, Improved performance of inverted planar perovskite solar cells with F4-TCNQ doped PEDOT: PSS hole transport layers, J. Mater. Chem. A 5 (2017), https://doi.org/10.1039/C6TA10212C.

[38] S. Li, Y.L. Cao, W.H. Li, Z.S. Bo, A brief review of hole transporting materials commonly used in perovskite solar cells, Rare Met. 40 (2021) 2712–2729, https://doi.org/10.1007/s12598-020-01691-z.

[39] H. Sharma, V.K. Verma, R.C. Singh, P.K. Singh, A. Basak, Numerical analysis of high-efficiency CH3NH3PbI3 perovskite solar cell with PEDOT: PSS hole transport material using SCAPS 1D simulator, J. Electron. Mater. 52 (2023) 4338–4350, https://doi.org/10.1007/s11664-023-10257-5.

[40] M.A. Nalianya, C. Awino, H. Barasa, V. Odari, F. Gaitho, B. Omogo, M. Mageto, Numerical study of lead free CsSn0. 5Ge0. 5I3 perovskite solar cell by SCAPS-1D, Optik 248 (2021) 168060, https://doi.org/10.1016/j.ijleo.2021.168060.

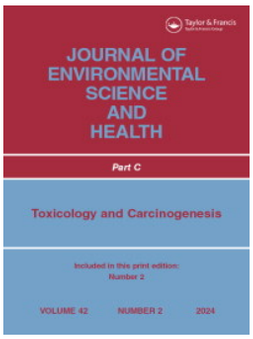
[41] N. Marinova, W. Tress, R. Humphry-Baker, M.I. Dar, V. Bojinov, S.M. Zakeeruddin, M.K. Nazeeruddin, M. Grätzel, Light harvesting and charge recombination in CH3NH3PbI3 perovskite solar cells studied by hole transport layer thickness variation, ACS Nano 9 (2015) 4200–4209, https://doi.org/10.1021/acsnano.5b00447.

[42] F. Azri, A. Meftah, N. Sengouga, A. Meftah, Electron and hole transport layers optimization by numerical simulation of a perovskite solar cell, Sol. Energy 181 (2019) 372–378, https://doi.org/10.1016/j.solener.2019.02.017.

[43] B.L. Williams, V. Zardetto, B. Kniknie, M.A. Verheijen, W.M.M. Kessels, M. Creatore, The competing roles of i-ZnO in Cu(In, Ga)Se2 solar cells, Sol. Energy Mater. Sol. Cells 157 (2016) 798–807, https://doi.org/10.1016/j.solmat.2016.07.049.

[44] F. Paquin, J. Rivnay, A. Salleo, N. Stingelin, C. Silva-Acuña, Multi-phase microstructures drive exciton dissociation in neat semicrystalline polymeric

semiconductors, J Mater Chem C Mater. 3 (2015) 10715–10722, https://doi.org/10.1039/c5tc02043c.
[45] S.J. Yoo, J.H. Lee, J.H. Lee, J.J. Kim, Doping-concentration-dependent hole mobility in a ReO3 doped organic semiconductor of 4,4′,4″-tris(N-(2-naphthyl)-N-phenyl-amino) -triphenylamine, Appl. Phys. Lett. 102 (2013), https://doi.org/10.1063/1.4804141.
[46] F. Anwar, R. Mahbub, S.S. Satter, S.M. Ullah, Effect of different HTM layers and electrical parameters on ZnO nanorod-based lead-free perovskite solar cell for high-efficiency performance, Int. J. Photoenergy 2017 (2017), https://doi.org/10.1155/2017/9846310.
[47] L. Hao, T. Li, X. Ma, J. Wu, L. Qiao, X. Wu, G. Hou, H. Pei, X. Wang, X. Zhang, A tin-based perovskite solar cell with an inverted hole-free transport layer to achieve high energy conversion efficiency by SCAPS device simulation, Opt. Quant. Electron. 53 (2021), https://doi.org/10.1007/s11082-021-03175-5.
[48] W. McKinney, Data structures for statistical computing in python, InProceedings of the 9th Python in Science Conference 2010 Jun 28 (Vol. 445, No. 1, pp. 51-56). DOI: 10.25080/Majora-92bf1922-00a.
[49] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117, https://doi.org/10.1016/j.neunet.2014.09.003.
[50] M. Hasanzadeh Azar, S. Aynehband, H. Abdollahi, H. Alimohammadi, N. Rajabi, S. Angizi, V. Kamraninejad, R. Teimouri, R. Mohammadpour, A. Simchi, SCAPS Empowered Machine Learning Modelling of Perovskite Solar Cells: Predictive Design of Active Layer and Hole Transport Materials, Photonics. 10 (2023). DOI: 10.3390/photonics10030271.



Taylor & Francis
Taylor & Francis Group

Journal of Environmental Science and Health, Part C
Toxicology and Carcinogenesis

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/lesc21

Primary Microplastics in the Ecosystem: Ecological Effects, Risks, and Comprehensive Perspectives on Toxicology and Detection Methods

Pooja Singh, Gunjan Varshney & Raminder Kaur

To cite this article: Pooja Singh, Gunjan Varshney & Raminder Kaur (05 Jul 2024): Primary Microplastics in the Ecosystem: Ecological Effects, Risks, and Comprehensive Perspectives on Toxicology and Detection Methods, Journal of Environmental Science and Health, Part C, DOI: 10.1080/26896583.2024.2370715

To link to this article: https://doi.org/10.1080/26896583.2024.2370715

Published online: 05 Jul 2024.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=lesc21

JOURNAL OF ENVIRONMENTAL SCIENCE AND HEALTH. PART C. TOXICOLOGY AND CARCINOGENESIS
https://doi.org/10.1080/26896583.2024.2370715

Taylor & Francis
Taylor & Francis Group

Check for updates

Primary Microplastics in the Ecosystem: Ecological Effects, Risks, and Comprehensive Perspectives on Toxicology and Detection Methods

Pooja Singh, Gunjan Varshney and Raminder Kaur

Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University (Formerly Delhi College of Engineering), Delhi, India

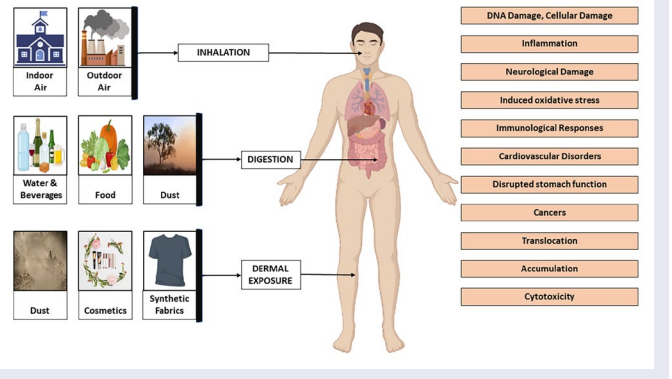
ABSTRACT

Recent discoveries of microplastics in cities, suburbs, and even remote locations, far from microplastic source regions, have raised the possibility of long-distance transmission of microplastics in many ecosystems. A little is known scientifically about the threat that it posed to the environment by microplastics. The problem's apparent size necessitates the rapid development of reliable scientific advice regarding the ecological risks of microplastics. These concerns are brought on by the lack of consistent sample and identification techniques, as well as the limited physical analysis and understanding of microplastic pollution. This review provides insight regarding some unaddressed issues about the occurrence, fate, movement, and impact of microplastics, in general, with special emphasis on primary microplastics. The approaches taken in the earlier investigations have been analyzed and different recommendations for future research have been suggested.

KEYWORDS

Primary microplastics; ecological risk; identification techniques; bioaccumulation; microplastic toxicity

GRAPHICAL ABSTRACT



CONTACT Raminder Kaur rkwalia14@gmail.com Discipline of Polymer Science and Chemical Technology, Department of Applied Chemistry, Delhi Technological University (Formerly Delhi College of Engineering), Delhi-110042, India.

© 2024 Taylor & Francis Group, LLC

1. Introduction: general aspects of microplastics

Plastics, regarded as a miracle material, have freed humanity from the socioeconomic constraints brought on by a scarcity of natural resources. Globally, fossil fuel feedstock accounts for 97–99% of plastics while the remaining 1–3% are derived from bio (plant) based. In the last few decades, the amount of plastic manufactured annually around the world has dramatically expanded, rising from 2 million tonnes in 1950 – 381 million tonnes in 2015.[1] After 2000, more than 50% of the total amount of manufactured plastic was introduced into the market. It is anticipated that their production would rise, even more, hitting approximately 600 million tonnes in 2025 (Figure 1). About 42% of the plastic produced is consumed by the packaging business, with the construction and building sector using the remaining 19%. Numerous "essentials" in our daily lives, such as cling film, face masks, medical supplies, coffee mugs, and shopping bags are made of single-use plastic (SUP). The recent Covid-19 outbreak increased the usage of SUPs, particularly gloves, facemasks, and personal protective equipment (PPE) kits, which added 3.5% more to the world's solid waste percentage. The US, India, and China are the largest producers of SUP, respectively.[2]

Plastics are ubiquitous in the environment due to their high manufacturing rates and longevity. Initially, larger plastic trash drew the attention of scientists and the general public. Tiny plastic particles, however, were found in the marine environment as early as the 1970s.[3] Microplastics (MPs) in the environment have gained more attention over the past several years and are now a prominent field of study. MPs are typically characterized as plastic particles that are less than 5 mm in size and have a wide range of shapes (including fibers, spheres, and fragments), sizes, colors, and compositions. Further, MPs can be categorized into primary and secondary regardless of their sizes. For application, for instance, in medical equipment and in personal care items like cosmetics and soaps, primary

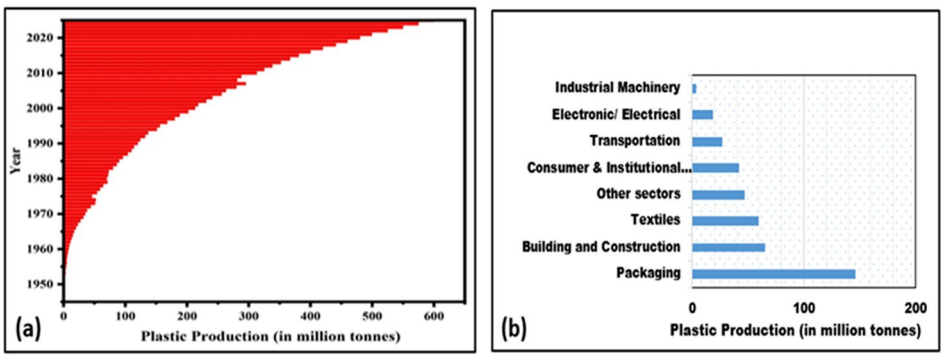


Figure 1. (a) Plastic production till 2025 and (b) primary plastic production by industrial sector, 2015.

MPs are found in microscopic sizes (fibers, microbeads).[4] Large plastic items, when in contact with the environment, physical, chemical, and biological factors cause the plastic to break down into tiny fragments, creating secondary MPs, including mulch breakage debris and tire-wear debris, among others. Terrestrial and aquatic ecosystems are found to contain enormous and diverse MPs. As MPs are difficult to degrade, they remain in the environment for a long time, which results in significant and harmful environmental pollution.[5]

This review addresses substantial knowledge gaps about the origin, destiny, migration, and effects of MPs, emphasising on primary MPs, which are becoming more widely acknowledged as a severe environmental threat. This study aims to provide a comprehensive understanding of the widespread presence of MPs in both terrestrial and aquatic ecosystems, their behavior, after entering the environment, as well as their potential negative effects on human and environmental health by integrating previous research findings. The goal of the review is to facilitate the development of evidence-based policies and interventions targeted at reducing the detrimental effects of MPs pollution by clarifying the sources, patterns of distribution, and ecological implications of primary MPs. Additionally, it offers guidelines for establishing effective risk management approaches for MPs by highlighting crucial areas for future research. This review seeks to give a thorough understanding of primary MPs and inspire the development of effective mitigation techniques, through a perspective, on which a relatively limited research is available.

2. Occurrence and distribution of MPs

Industrial, domestic and coastal activities are the prime anthropogenic actions that have resulted in the presence of these harmful plastic particles in the ecosystem (aquatic and terrestrial). In addition, the fragmentation of huge plastic garbage and residential runoff, comprising MPs fragments and microbeads (used in cosmetics and other consumer goods), are the major contributors to the entry of MPs into the aquatic ecosystem (Figure 2).[6] The air-blasted resin pellets and powders that are released by the plastic manufacturing sectors also end up contaminating the aquatic environment. MPs pollution in the marine ecosystem is also caused by coastal activities like aqua tourism, marine industries, and fishing.[7] MPs pollution in soils is mostly attributed to sewage irrigation, plastic trash, agricultural facilities (such as agricultural compost), surface runoff, fertilizer use and atmospheric deposition.[8] In addition, MPs are produced during the stacking, disposal, or burning of plastic trash. Other sources of MPs in the environment include clothes, industrial dust emissions, housing infrastructure, and other daily life activities. MPs can be suspended and transported

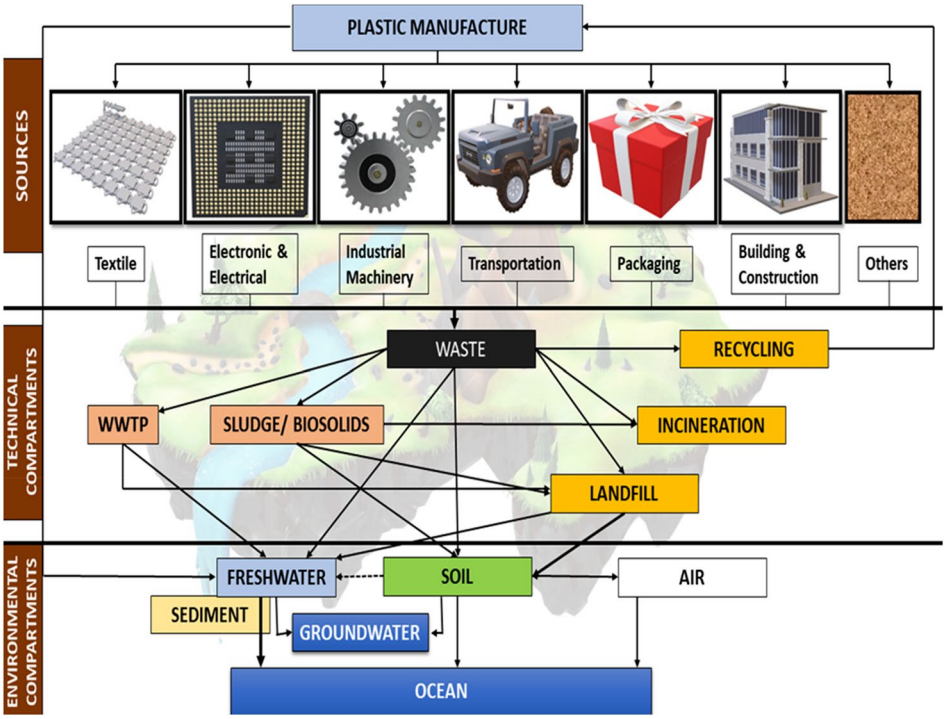


Figure 2. Occurrence and distribution of plastics into different environmental compartments.

by air in the atmosphere since they are naturally low in weight.[9] MPs temporal and spatial distribution across the soil, aquatic systems, and the atmosphere is controlled by the transit of those materials, as these can migrate between multiple ecological divisions. For instance, through human activity, wind erosion, and runoff, MPs in the soil can also be carried into neighboring environmental matrices including the water and air.[10] Their movement within the soil is influenced by a number of aspects, including soil texture, cracking, biotic transport (soil biodiversity and soil community composition), and aggregation.[11]

The vertical transportation of MPs in the soil is facilitated by agricultural methods, soil fissures, root growth, and soil pores. Organisms have also been observed to be involved in the mobility and distribution of MPs in soil.[8] Additionally, it's possible for MPs in the soil to get into groundwater. Pollution from MPs in the environment may end up in the oceans and on land. Numerous elements, including wind direction, wind speed, particle density, and rain, have an impact on the transportation of MPs and their fallout or deposition in the atmosphere.[12] While some MPs settle through dry and wet deposition processes, others remain suspended in the atmosphere or eventually settle into aquatic and terrestrial ecosystems.[13] Further, under intense and ongoing hydrodynamic events like waves,

rainstorms, ocean currents, monsoons, and tides, the aquatic environment is a dynamic region, and thus, MPs can travel between the marine environment and the land.[11] The fundamental characteristics of MPs, such as shape, density, color and size, can have significant impact on their transportation.[14]

3. Classification and sources of MPs

MPs are classified into two main types based on their origin and manufacturing processes: primary and secondary. Primary MPs are intentionally manufactured particles, such as microbeads in cosmetics, resin pellets used in plastic production, and fibers from textiles. MPs in the micro-meter size are specifically manufactured for use in industrial abrasives for sandblasting, such as polyester or acrylic beads, plastic pre-production pellets ('nurdles'), or personal care products like exfoliating agents in creams and cleansers comprising polyethylene 'microbeads'. Primary MPs particles are most likely to reach wastewater treatment streams, after being washed down by industrial or domestic drainage systems. In spite of the capability of a few sewage treatment works to evacuate up to 99.9% the absolute number of particles entering the system may still permit a noteworthy number to bypass filtration frameworks, and be discharged into the freshwater environment with wastes.[15]

Secondary MPs are created as a consequence of the fragmentation of the macro and meso-plastic litter, resulting from the degradation and disintegration of larger plastic items, such as bottles, bags, and packaging materials owed to environmental factors like UV radiations, mechanical abrasion, and microbial activity.[16] High temperatures and UV radiation can induce chemical changes in plastics, causing them to become brittle and therefore more prone to the fragmentation. Fragmentation allows greater surface area and particle count per unit mass. The contributing factors of fragmentation in marine waters are both wave action and sunlight. Plastic fragmentation is considered to occur readily on land, particularly at the soil surface, as a consequence of exposure to ultraviolet radiation from the sun, assisted also by temperature variations that are usually greater than those found in sea water. Despite the fact that microfibres are secondary particles, they will be released into the environment alongside primary MPs, via sewage effluents as well as sludge disposal. As a result, based on similar release routes, the fate and transfer of such fibers could be more closely linked with that of primary MPs. Understanding the distinction between the primary and the secondary sources is crucial for comprehensively assessing the impact of MPs on the environment and human health as depicted in Figure 3.

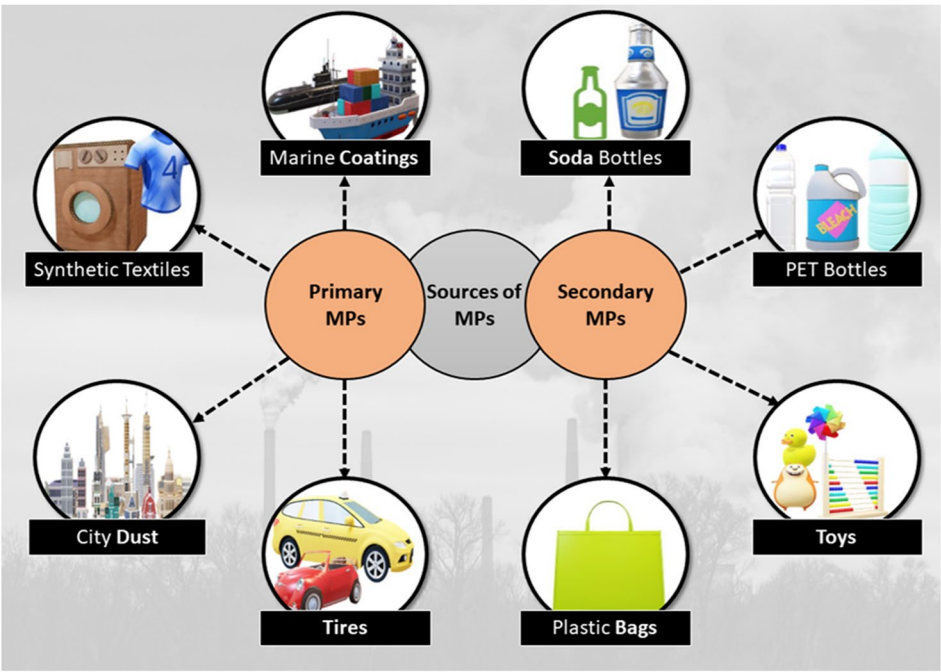


Figure 3. Primary and secondary sources of MPs in the environment.

3.1. Primary MPs

MPs particles are employed as exfoliants in some personal care product segments, such as hand cleaners, facial cleaners, and toothpastes.[17] Polyolefin particles exhibiting a size of 74-420 µm and an amorphous frame, with no sharp edges were characterized as suitable for use as exfoliants in a US patent for skin cleaners containing plastic microparticles.[18] Polyolefins used in them include polyethylene (PE), polypropylene (PP), and polystyrene (PS). Scrubs have been shown to contain PE microbeads in concentrations ranging from 0.4 to 10.6%, with sizes ranging from 40 to 1000 µm[17]. According to a Cosmetics Europe survey, PE accounted for 93% of the MPs used in skin cleaner products in these countries in 2012. MPs are also utilized in medical settings, such as in dental tooth polish and also as delivery systems for pharmaceutical active ingredients.[19] MPs from medical and personal care products can be released into the environment post used via wastewater. MPs are also used in industrial abrasives used for air-blasting the paint off the metal surfaces and cleaning various types of engines, as well as drilling fluids for oil and gas exploration.[20] MPs made of poly allyl diglycol carbonate, acrylic, PS, melamine, and polyester (PES) are among the materials found in industrial abrasives.[21] These may enter the environment if used in open systems or improperly disposed off. Another significant source of primary MPs is the raw material utilized to make plastic products (pre-production plastics), specifically plastic resin

flakes or pellets and fluff or plastic powder that can enter the environment as a result of improper handlings, such as through run-off from processing plants or accidental loss during transit. Furthermore, degranulate generated during the recycling of plastic and remnants from plastic processing factories can eventually reach the aquatic or terrestrial ecosystems.

Glitter, a form of MPs, comprised of tiny, flat, shiny particles, is commonly used in household items, cosmetics, and arts and crafts. It is typically made of Mylar^{TM}, a type of polyester film known as BoPET (biaxially-oriented polyethylene terephthalate), which has a melting point of 260 °C and a real density of 1.38 g/cm^3.[22] These particles are water-insoluble and often coated with metal to enhance the reflectivity. While plastic glitters come in a wide range of colors, shapes, and sizes; their environmental impact is a growing concern due to their potential to harm the ecosystems, as they can easily enter the environment and pose risks to life and human health. Though, some manufacturers have introduced biodegradable glitter alternatives made from plant-derived materials; despite this, the widespread use of plastic glitters continues, to raise the environmental and health concerns. Therefore, it is essential to address the ecological and health implications of glitter as a primary MPs, and to explore sustainable alternatives to mitigate its environmental impacts.[23]

3.2. Secondary MPs

Secondary MPs are produced when bigger plastic materials break down into smaller pieces. The four human activities responsible for the plastic entery into the environment are thought to be general littering, the dumping of plastic waste, loss from poorly managed landfill sites, and waste collection.[24] Additionally, waste from the recycling facilities is lost due to wind contributes significantly toward the MP polluton.[25] A huge portion of the plastics, that is produced and used for packaging of goods, with a limited shelf life is habitually dumped in the landfills and is often covered periodically with soil or synthetic material to prevent the trash from being blown away. During natural calamities like strong seas, tsunamis, and hurricanes, substantial amounts of these plastic particles may be released into the ocean.[25] Another significant source of MPs found in the environment is low-density polyethylene (LDPE) films, which are employed in enormous quantities to raise soil temperature, protect crops, suppress weeds, and hold irrigation water in the soil (a process known as "plastic mulching"). The particles from these thin plastic foils could wind up in the soil if they crumble. Additionaly, a significant culprit of MPs release in the fibers released after washing PES blankets, fleeces, and shirts in home washing machines. During a single wash, the tested PES fleece released over 1900 fibers in total. Also, synthetic textile fibers are discharged into the air during routine use or tumble

drying.[26] There are numerous other places from where MPs can enter the environment, such as car tyre abrasion and application of paints & protective coating. Abrasion from other plastic materials, such as household plastics, results in the release of MPs as well.[27] Material lost or discarded from fishing boats, aquaculture facilities, merchant ships, recreational boats, offshore oil or gas platforms, and during military operations are examples of ocean-based sources of marine litter. The dominant form in the environment is still unknown, but it is likely location-dependent; for example, the primary MPs may be more significant in areas close to wastewater discharge sources than secondary MPs, which may predominate in open waters. Table 1 provides the spatial and temporal pattern in the abundance of Primary MPs.

4. Characteristics of MPs

MPs pollution of the environment is now recognized as a risk to ecosystem and researchers have begun to address this field as an issue of importance in recent years, after recognizing the comparative lack of studies on MPs. The effects of MPs on soil ecosystems (both below and above ground) are almost unexplored. Furthermore, the effects of MPs on terrestrial systems vary depending on the type of MPs. It has been demonstrated that the physical and chemical characteristics of MPs influence their ability to bind pollutants.[51]

4.1. Physical properties of MPs

4.1.1. Size

The sampling and analytical techniques are frequently used to identify the sizes of micro/nano-plastics. While MPs come in a variety of sizes, the most prevalent range from 1 µm to 5 mm, nano-plastics (NPs) are approximately <100 nm in size.[52] The research on MPs size has gradually shrunk to less than 2000 µm, due to the progress made in the qualitative and quantitative approaches for MPs. The presence of smaller MPs in the atmosphere warrants additional investigation, as evidenced by the abundance of fluorescent non-fibrous particles <20 µm and smaller fibers that were found in the total atmospheric deposition in downtown London.[53] These investigations demonstrated that MPs, which can be inhaled and consumed, make up a non-negligible portion of the atmospheric particulate matter. Additionally, the research suggests that inhaling MPs could be harmful to human health.

4.1.2. Shape

The most frequent morphologies in atmospheric MPs are fragmented and fibrous. In addition, there are other shapes such as spheres, flakes, and

Table 1. Spatial and temporal pattern in the abundance of primary MPs.

S. No.	Country	Description	Ref.
1.	Mainland China	In 2015, one-sixth of the 737.29 Gg primary MPs waste from mainland China made its way out into the water. Tire dust and synthetic fiber accounted for the largest percentage of this trash, with 53.91 and 28.77% of the total, respectively.	28
2.	Netherlands	In STP effluent, the total estimated concentrations of primary MPs from consumer items are 0.2, 2.7, and 66 μ/L in three different scenarios. The three product categories under consideration, namely paint and coatings, cleaning agents, and cosmetics and personal care items, are all significant contributors.	29
3.	East Asia (marine medaka *Oryzias melastigma*.	Marine medaka embryos exposed to primary MPs experienced a variety of adverse outcomes, such as heart rate alterations, morphological irregularities, and various forms of deformity.	30
4.	Sweden	Major kind of detected MPs are fibers and fragments: Influent MPs concentration 15MP L^{-1} Effluent MPs concentration 8.3MP L^{-1} and Daily MPs discharge 3.6×10^{-4} MP d^{-1b}	31
5.	Finland (Baltic Sea)	Major kinds of detected MPs are synthetic particle and textile fiber: Influent MPs concentration 430, 180MP L^{-1} Effluent MPs concentration 9, 5MP L^{-1} and Daily MPs discharge 3.7×10^{9} MP d^{-1}	32
6.	Scotland (River Clyde, Glasgow)	Major type of detected MPs are fibers (polyamide, polyester, acrylic): Influent MPs concentration 16MP L^{-1} Effluent MPs concentration 0.3MP L^{-1} and Daily MPs discharge 6.5×10^{7} MP d^{-1}	33
7.	Germany (Lower Saxony)	Major type of detected MPs are fibers (polyethylene): Daily MPs discharge 4.2×10^{4}–1.2×10^{7} MP d^{-1}	34
8.	Australia (Sydney)	Major type of detected MPs are fibers (polyethylene): Influent MPs concentration 2MP L^{-1} Effluent MPs concentration 0.3,0.2MP L^{-1} and Daily MPs discharge 3.6×10^{6}–1×10^{7} MP d^{-1}	35
9.	Canada (Vancouver)	Major type of detected MPs are fibers and fragments: Influent MPs concentration 14MP L^{-1} Effluent MPs concentration 1MP L^{-1} and Daily MPs discharge 10.6×10^{9}–19.9×10^{9} MP d^{-1}	36
10.	Turkey (Seyhan and Yüreğir WWTP in Adana city)	Major type of detected MPs are fibers (polyester): Influent MPs concentration 4.8×10^{6} MP per day; 2.0×10^{6} MP per day Effluent MPs concentration 1.2×10^{6} MP per day; 0.35×10^{6} MP per day and Daily MPs discharge 2.2×10^{5}–1.5×10^{6} MP d^{-1}	37
11.	USA (Charleston Harbor, South Carolina)	Major type of detected MPs are fibers, microbeads: Daily MPs discharge 5×10^{8}–1×10^{9} MP d^{-1}	38
12.	China (WWTP located in Wuxi, Jiangsu Province, Eastern China)	Major type of detected MPs are fibers and fragments: Influent MPs concentration 4MP L^{-1} Effluent MPs concentration 0.1, 0.1MP L^{-1} and Daily MPs discharge 6.5×10^{6}–3.5×10^{6} MP d^{-1}	39
13.	Italy (Northern Italy)	Major type of detected MPs are fibers (polyamide and polyesters: Influent MPs concentration 3MP L^{-1} Effluent MPs concentration 0.4MP L^{-1} and Daily MPs discharge 10.6×10^{8} MP d^{-1}	40

(*Continued*)

Table 1. Continued.

S. No.	Country	Description	Ref.
14.	Netherlands (Dutch River Delta and Amsterdam Canals, Waste water treatment plant, North Sea Sediments and biota)	Major type of detected MPs are fibers: Influent MPs concentration 68–910MP L^{-1} Effluent MPs concentration 52MP L^{-1} and Daily MPs discharge 7.5×10^{8}–4.3×10^{10} MP d^{-1}	41
15.	Russia (Central WWTP of Vodokanal of St. Petersburg)	Major type of detected MPs are fibers, synthetic particles: Influent MPs concentration 467, 160MP L^{-1} Effluent MPs concentration 16, 7MP L^{-1}	42
16.	France (River Seine)	Major type of detected MPs are fibers: Influent MPs concentration 260–320MP L^{-1} Effluent MPs concentration 14–50MP L^{-1} and Daily MPs discharge 8.4×10^{9} MP d^{-1}	43
17.	South Korea (STF of M-city, Y-city, and S-city)	Major type of detected MPs are fibers and fragments: Influent MPs concentration 30, 17, 14MP L^{-1} Effluent MPs concentration 0.4, 0.1, 0.3MP L^{-1}	44
18.	Denmark (10 largest WWTP)	Major type of detected MPs are fibers and synthetic particles: Influent MPs concentration 7216MP L^{-1} Effluent MPs concentration 54MP L^{-1}	45
19.	Poland (Southern Poland)	Major type of detected MPs are fibers: Influent MPs concentration 19–552MP L^{-1} Effluent MPs concentration 0.028–0.96MP L^{-1}	46
20.	U.S. (Florida)	Prove that MPs have contaminated the tissues of intertidal marine predatory snails: 256 MPs were extracted from snails ($n=60$), of which 93% were microfibers and 7% were microfragments. Furthermore, 67 MPs (8.375 MP/L) of which 97% were microfibers and 3% were microfragments were extracted from 8 L of seawater.	47
21.	South Australia (Bellevue Heights, 5050)	MPs can be found in gardens as a result of using plastic landscaping fabrics in the mulch bed or by leaving plastic bubble wrap to be combined with mulches. They were initially quite huge, measuring more than 5 mm. However due to abrasion, weathering, or natural decay, MPs are freed from the garden after seven years.	48
22.	Rome, Italy (Ospedale Fatebenefratelli Isola Tiberina)	Detection of MPs in human breastmilk Major type of detected MPs are nitrocellulose, polyethylene, polyvinyl chloride, polypropylene, chlorinated polyethylene, polyvinyl alcohol. Breastfeeding moms may expose their infants to toxins from beverages, foods and personal care products, which could have a harmful impact.	49
23.	U.S.(Orontes River)	Major type of detected MPs are fibers and fragments: Influent MPs concentration 57.2MP L^{-1} Effluent MPs concentration 2.1MP L^{-1}[a]	50

[a]MP L^{-1}: Millions Particles per litre
[b]MP d^{-1}: Million Particles per day

films. MPs have distinct physical characteristics depending on their shape. Fibers are typically very thin, flakes are typically thick and asymmetrical, and films are typically thin and flexible. Overall, the primary MPs morphology at the time of manufacture, surface deterioration and erosion processes, and residence time in the environment all affect the form of MPs in the atmosphere.[54]

4.1.3. Color

One of the key aspects for the identification of plastic debris and potential contaminants during sample preparation is color. There have been reports of MPs/NPs in a variety of colors, including green, red, blue, orange, grey, yellow, white, brown, off-white, and tan. For the visual evaluation of MPs particles in the atmosphere, color is generally a helpful parameter. While low-density polyethylene (LDPE) has been given opaque colors, polypropylene has been clear and transparent. When reporting and interpreting results, it is important to take into account that MPs discoloration can happen during sample preparation (H_2O_2-oxidative digestion) and weathering.[55] Like natural foods, brightly colored NPs particles are frequently consumed by marine species. MPs in the colors black, blue, green, and red are primarily consumed by copepods, zooplankton, fish larvae, and euphausiid.[56]

4.1.4. Crystallinity

Plastics density, permeability, and swelling characteristics are affected by the organized structural links known as crystallinity.[57] As MPs/NPs spend more time in their surroundings, their crystallinity changes. The polymer's amorphous area breaking down encourages overall crystallinity and shrinks MPs/NPs size.[11] Crystallite materialization may change MPs/NPs from their counterparts and change their toxicity. It ultimately affects the surface area, particle shape, size, and chemical characteristics such as the adsorption of contaminants and additives, which in turn affects the ingestion rate.[58]

4.1.5. Surface properties

Surface chemistry and surface area are the two main components of surface property. Particles in the nanoscale have a significant impact because MPs/NPs surface area grows as particle size decreases.[58] The deterioration of the plastic surface by photo- and oxidative processes results in the formation of new functional groups when OH radicals, O and N oxides, and other photo-generated radicals are interacted with.[59] Due to the disruption of the plastic's surface area by these pathways—chemical leaching, creation of angular-shaped NPs from primary microbeads, and ingestion—microscopic particles are liberated.[57] Furthermore, because the microbial community uses the oxygenated binding sites, surface chemistry has a substantial impact on particle-biota interactions.[60]

4.2. Chemical properties of MPs

MPs comprise additives (plasticizers, antioxidants), polymers, dyes, and pollutants that get adsorbed on the surface. When plastics are produced, utilized, and deteriorate, these chemicals are easily discharged into the environment.[61,62]

4.2.1. Chemical additives

To increase the product's efficiency, additives are added during the production of plastic to achieve the desired color, and transparency. These additions increase the polymer's resistance to degradation by biological (bacteria and fungi), mechanical, physical (light radiation, ozone, and temperature) and electrical agents.[63] Additives include pigments, ultraviolet stabilizers, flame retardants, lubricants, and plasticizers as well as inert or reinforcing fillers. Additional ingredients include wood, clay or linen, graphite, cellulose pulp, glass fibers, jute, rock flour, cotton flakes and kaolin.[64] Low molecular weight additives embedded in the polymer matrix are brittle and can escape out into the neighboring water bodies (Polyethylene terephthalate (PET) oligomers from food trays and bottles; flame retardants from electronic items; nonylphenol from food contact materials; Lead (Pb) from un-plasticized PVC pipes; and Antimony (Sb) from PET water bottles).[59] The majority of MPs and NPs are additives and fundamental polymeric substances made from plastics and chemicals absorbed from the environment and are linked to plastic toxicity (i.e., monomers, plasticizers, solvents, additives-dyes, and catalysts).[60] These substances disturb the endocrine system, when ingested or inhaled, leading to metabolic diseases, hormonal imbalance, neuro-developmental abnormalities, reproductive issues and asthma. The impact of various additives in plastics on human health is given in Table 2.

4.2.2. Chemical structure

Higher sorption capacity may result from a chemical structure with a wide surface area or high porosity. For instance, PET had the highest sorption rate, while PE had the largest sorption capacity. The PE, with a higher sorption capacity, leads to more pollutant diffusion, due to its bigger surface area, wider space between polymer chains, and free volume.[67] As opposed to PE, PET has a smaller surface area and a glassy polymeric structure, which do not facilitate diffusion into the material and so have a faster sorption rate.

4.2.3. Chemical interactions

Pollutant affinity for MPs varies due to chemical interactions such as hydrogen bonding, electrostatic interaction, π–π interaction, hydrophobic interaction, or van der Waals force.[68] For example, because of the amide groups, hydrogen bonding between polyamide (PA) and antibiotics caused strong affinity.[69] The difference in the surface charges between amine-modified polystyrene (PS-NH_2) and carboxyl-functionalized polystyrene (PS-COOH), caused embryonic apoptosis, oxidative stress, cell death, and

Table 2. The impact of various additives in plastics on human health.[11,65,66]

	Biocides	Stabilizers, antioxidants, and organic pigments	Plasticizers	Flame retardants
Additives	• Heavy metals- Antimony, copper, mercury, arsenic, tin • Butyltin trichloride • Tetrabutyltin • Arsenic trioxide	• (TGIC) Triglycidylisocyanurate • (TNPP) Tris-nonyl-phenyl phosphate • (BHA) 2-t-butyl-4 hydroxyanisole • Irganox 1010	• (BBP) Butyl benzyl phthalate • (MDA) Formaldehyde, 4,4′-methylenedianiline • (DIHP) Diisoheptylphthalate • (DCHP) Chlorinated paraffins, Dicyclohexyl phthalate	• TCPP) Tris(2-chlorisopropyl)phosphate • Boric acid • (BDE 47) 2,2′,4,4′-Tetrabromodiphenyl ether • Heavy metals—antimony, bromine, aluminum, zinc, bromine
Effect on human health	• Breast cancer • Metabolic and mental disorders • Kidney disease • Neuronal toxicity • Cardiovascular disease	• DNA methylation • Miscarriages • Hypertension • Brain damage • Anemia	• Neuro-degenerative disorder • Neuronal toxicity • Metabolic and mental disorders • Kidney disease	• Genotoxicity • DNA methylation • Apoptosis • Osteomalacia and bone fractures

disruption of the cell membrane in the embryos of the sea urchin Paracentrotus lividus.

5. Bioaccumulation of MPs

5.1. Transfer mode of MPs in human body

MPs enter the human body through three main routes: ingestion, inhalation and skin contact (Figure 4).[70] Due to their prevalence in the water supplies and food chain, MPs are easily consumed.[71] MPs are not able to penetrate the epidermal membrane being too thin, but they might still enter through hair follicles, wounds, or sweat glands. As they can easily enter the environment and pose risks to marine life and human health. The fact is that human body is exposed to MPs through all three channels, although environmental and seafood particles pose the highest danger of absolute exposure.

5.1.1. Gastric exposure

Recent investigations on MPs exposure and toxicity have shown that ingesting plastic particles is the main way that people are exposed to.[72] There are studies demonstrating that MPs are being consumed through food and drink, but there aren't any explicitly examining the toxicity of NPs in individuals.[73] Human feces samples were initially examined, and the results revealed that plastic particles were being ejected, supporting the theory that people are consuming these particles through water and food. These observations, along with studies on ingested uptake in environmental models, unequivocally demonstrate that MPs and NPs consumption by humans will be pervasive.[74] It would be vital to investigate their pathway

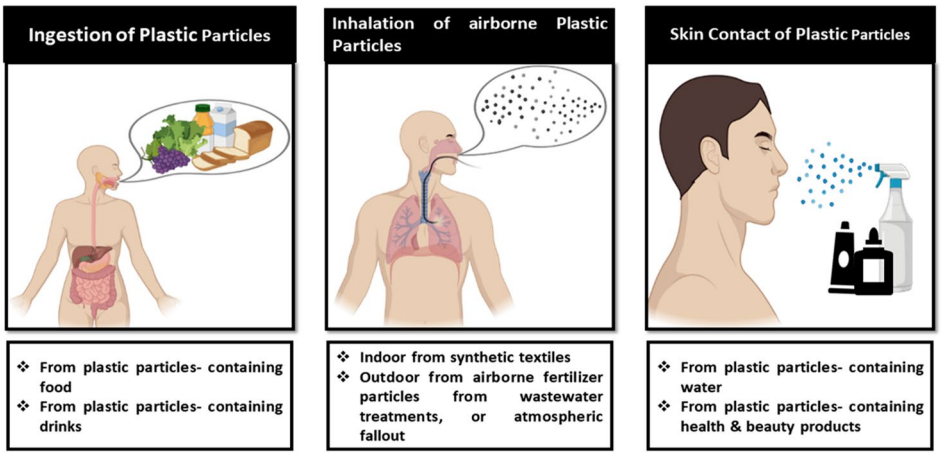


Figure 4. Three main routes by which plastic particles enter the body: Inhalation, ingestion, and skin contact.

via the GI tract and identify whether these particles translocate across the gut epithelia or persist in the gut lumen. MPs ability to penetrate at the paracellular level is uncertain since the pores at the tight junction channels i.e., relevant have a maximum functional size of about 1.5 nm. It is possible that they will enter via lymphatic tissue, and it is more likely that they enter through endocytosis or phagocytosis and invade the Peyer's patches' microfold (M) cells.[75]

5.1.2. Pulmonary exposure

Inhalation is the second most likely way for a human to come into contact with NPs. Indoor spaces have airborne plastic particles, mostly from synthetic textiles, which can cause occupational exposure or unintentional inhalation.[76] Exposure in outdoor areas may occur by inhaling contaminated aerosols from the ocean's waves or airborne fertilizer particles from dry wastewater treatment processes. The lung's alveolar surface area is substantial—about 150 m^2—and its tissue barrier is very thin—less than 1 μm. Nanoparticles can distribute throughout the entire human body because this barrier is permeable enough to pass through and into the capillary blood system.[72]

5.1.3. Dermal exposure

MPs/NPs are also commonly found in beauty and health products, especially in body and facial cleansers that are used directly on the skin.[77] Although there isn't enough evidence to draw firm conclusions on the impacts of nanocarriers, skin penetration depends heavily on the size of the particles and the health of the skin. Only one study describes that the synthetic textile causes the penetration of nanoparticles on the epidermal barrier at a very minute quantity. The stratum corneum, the skin's outermost layer, serves as a barrier to protect the skin from injury, microbial agents, and chemicals. Corneocytes comprise the stratum corneum, which is encased in lamellae of hydrophilic lipids which include long-chain free fatty acids, cholesterol, and ceramide. The skin could be exposed to plastic particles by contact with contaminated water or through the use of beauty and health products.[75] Plastic particles might enter the body through sweat glands, skin wounds, or hair follicles because MPs and NPs are hydrophobic, making it unlikely to pass through the stratum corneum and be absorbed.

5.2. Transfer in the bodies of animals (terrestrial and aquatic)

5.2.1. Terrestrial animals

MPs are ingested orally, where they enter the body, travel through the digestive system and is subsequently eliminated. However, a small number

of MPs can persist for days in the intestine of organsms.[78] Cells can absorb MPs in the intestine through endocytotic processes.[79] MPs that enter the intestines have also been demonstrated to induce inflammation and mechanical damage. Because of this damage to the intestinal epithelium, MPs can easily spread to other tissues and organs of organisms or enter their bloodstream through intestinal capillaries. Recently, it was shown that NPs can permeate the mucosa and muscularis layers of the intestine, pass beyond the gut epithelium, and then be transferred into other organs (such as the liver) by blood circulation.[80] This was done using an *ex vivo* salmon (*Salmo salmar*) gut sac exposure system.[81] Additionally, MPs can physically break up into smaller pieces as they go through the digestive system, which helps to facilitate their translocation.[82] An inhalation investigation in maternal rats also revealed that 20 nm NPs translocated to the fetal liver, brain, heart, kidney, and lungs by intratracheal instillation in addition to placenta, heart, and spleen during gestation.[83]

5.2.2. Aquatic animals

MPs in aquatic animals' gill canals could be directly injected, translocated to the gastrointestinal tract, or intercepted and then conveyed to the capillaries *via* the gill filaments.[84] Exposure to MPs caused several types of damage to fish gills, such as the lifting of epithelial, inner opercular membrane swelling, and aneurysms in secondary lamellae. Another potential pathway for MPs entrance into animal bodies is through the epidermis. By adhering to animals' damaged skin during the skin's healing phase, MPs can infiltrate muscle tissue. According to certain reports, some plastic that was wrapped around sea turtles eventually grew into their muscles.[85]

6. Fate of MPs

MPs appear to have no limits in the environment, and they are now pervasive throughout all ecosystems in the world (Figure 5). Plastic waste that has been released into the environment can be carried by the wind, swept from the land to surface waterways during rain, particularly with stormwater runoff, and transported in both freshwater and seawater. Large rivers are expected to transport a significant amount of micro- and macro-plastics into the oceans, although there aren't many studies to support this hypothesis. Over 50% of the world's population lives within 80 kilometers of the sea, thus, wind transport of macro-plastics to the oceans may be considerable. Very small MPs, which could be mobilized from open landfills, for example, may potentially be appropriate for airborne transport.

Depending on the density, plastics will either hover or submerge in water. Plastics having a density less than freshwater (about 1000 kg/m^3) or seawater (about 1030 kg/m^3) will be buoyant. The majority of consumer plastics, including PE, are buoyant in saltwater (Table 3). Plastics are typically colonized by a variety of species in aquatic environments. By increasing the density, this fouling can cause formerly buoyant objects to sink beneath the water's surface. The density of a plastic object may decrease if it becomes fouled by other species, causing it to resurface in the water. The specific density of MPs may also vary due to erosion, and wind-driven mixing of the upper water layer may cause previously buoyant MPs to sink. Depending on ocean currents, local winds, and shoreline geography, plastics can be transported over great distances. Beaches and oceanic gyres become overrun with buoyant plastic trash, and benthonic debris which occurs on the sea floor in the region with little circulation. Because smaller particles at the water's surface are less wind-exposed than larger ones, the mobility of MPs in surface waters may vary from that of macro-plastics.[87] Animals activity and aggregation of particles may also have an impact on the distribution of MPs. MPs can be covered by silt on beaches and in subtidal sediments, where they can be found at great depths. The movement of MPs throughout the terrestrial environment is still mostly unknown. It was established that the assimilation of micro- and macro-plastic particles into the soil is facilitated by terrestrial organisms like moles and earthworms. Plastic's photo-oxidative degradation is

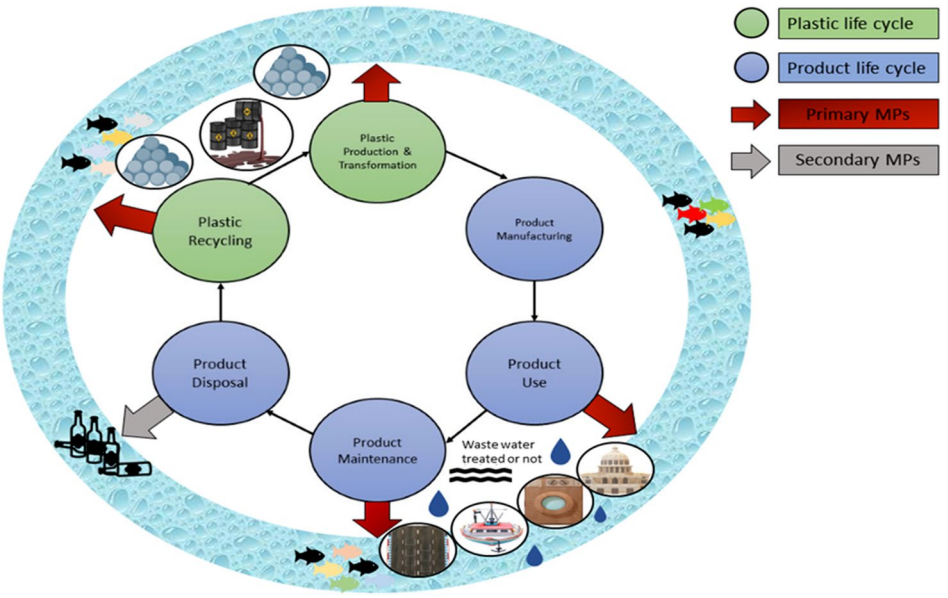
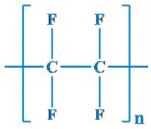
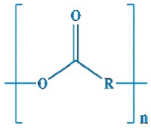
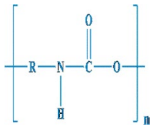
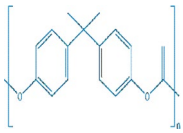
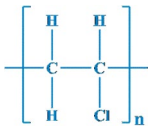
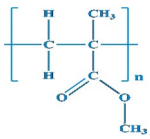
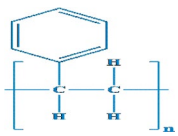
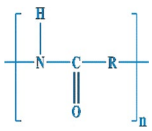
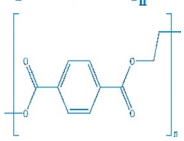
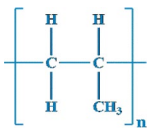


Figure 5. Plastic and plastic product lifecycle.

Table 3. Plastic material densities that are frequently observed in aquatic ecosystems.[86]

Abbreviation	Structure	Plastic class	Density (kg/m^3)
PTFE		Polytetrafluoroethylene	2100–2300
PES		Polyester	1240–2300
–		Alkyd	1240–2100
PU		Polyurethane	1200
PC		Polycarbonate	1200–1220
PVC		Polyvinylchloride	1160–1580
PMMA		Polymethyl methacrylate (acrylic)	1090–1200
PS		Polystyrene	1040–1100
PA		Polyamide (nylon)	1020–1160
PET		Polyethylene terephthalate	960–1450
PP		Polypropylene	830–920
HDPE	–	High-density polyethylene	940–980
LDPE	–	Low-density polyethylene	890–930
EPS	–	Expanded polystyrene (Styrofoam)	10–40

somewhat effective on a beach surface, but very slowly for plastics buried in soil or beach silt, or deep ocean.[88]

7. Ecological risks of MPs

7.1. Effects of MPs on the aquatic ecosystem

Atmospheric MPs penetrate the aquatic ecosystems by both dry and wet deposition. Thus, atmospheric deposition is a significant cause of the occurrence of MPs in the aquatic environment. Due to their longevity, MPs are persistent in aquatic conditions. Some MPs, might sink to the ocean floor as a consequence of having a higher density than sea water, whereas other MPs that are less dense than seawater frequently float on the water. Organisms in freshwater and the ocean may consume MPs.

7.1.1. Effects on aquatic animals

Diverse aquatic organisms, including invertebrates, mammals, fish, shellfish, and fish, consume MPs that enter the aquatic ecosystem. Both active ingestion of seawater and passive ingestion of prey can expose mammals to MPs.[89] Because of the chemicals involved, MPs can have an impact on the gut microbiome of aquatic animals.[90] They may build up in the gastrointestinal tracts of aquatic animals, preventing them from ingesting and properly digesting food.[91] Aquatic animals may become poisonous or even die as a result of this. Ingestion of modest dosages of MPs does not always result in the rapid death of organisms. MPs can have a major toxicological impact on fish, including the introduction of oxidative stress and inflammation in the gut. Non-aquatic animals like penguins and seagulls have been known to accidentally consume MPs that float on the water's surface. The MPs wafting on the water are susceptible to intentional or unintentional feeding by zooplankton because they resemble their food in size and colour.[92] Microfibres, mostly generated from synthetic textiles, seem to have a higher possibility than other fibers to infiltrate the food chain, as a consequence of ther size and shape. Ther size permit them to be promptly consumed by the aquatic organisms and are more susceptible to becoming entwined in large lumps inside the gut, instigating blockages.[93] In addition to having a deleterious impact on the flora and fauna directly, MPs can concentrate on living beings, reach the food chain, progress to various trophic levels, and potentially pose a risk to human health.

7.1.2. Effects on aquatic microorganisms

MPs increase the amount of inorganic and organic carbon in the aquatic environment, which increases microbial activity[94] and can cause additional

CO_2 to be released into the aquatic environment resulting from changes in the CH_4 cycle. The MPs invasion altered the structure and composition of the bacterial communities and reduced the diversity and richness of the biofilm.[95] Food webs and carbon metabolism in aquatic environments are impacted by changes in the aquatic microbiome. Microorganisms prefer using MPs as novel substrates because of their buoyancy, hydrophobic surface, and capacity for long-distance transmission.[96] The colonization of pathogenic microbes on the surface of MPs may impact the transfer of deadly microbes into the aqueous environment.[97]

7.1.3. Effects on hydrophytes

MPs that float on the water's surface inhibit phytoplankton from absorbing light, which prohibits them from offering oxygen and food to aquatic organisms.[98] The architecture of phytoplankton communities can be drastically altered by high MPs concentrations.[99] Food webs and ecological processes in the aquatic environment will be impacted by the modifications to phytoplankton colonies. MPs also impair chlorella's capacity for the process of photosynthesis through oxidative stress and physical deterioration.[100] Algae could become more prevalent in some areas of the phytoplankton habitat. Because they are generally denser than water, common MPs gravitate into aquatic habitats. Submerged plants will experience oxidative stress, antioxidant defenses, and chlorophyll fluorescence characteristics because of these MPs, which will result in decreased plant height.[101] Plants submerged in water may be exposed to MPs through their leaves or roots.[102] The presence of microfibrils in the cell wall provided evidence that MPs had penetrated the plant.

7.2. Effects of MPs on the terrestrial ecosystem

MPs are introduced into the terrestrial environment through littering and the dumping of sewer sludge to land. MPs have the capability to penetrate the terrestrial environment through atmospheric transport and deposition. Studies report that both dry and wet deposition are ways in which atmospheric MPs penetrate the terrestrial environment.[103] Tyre plastic may escape into the ground from the road.[104] MPs from the atmosphere that enter the terrestrial environment either linger in the soil for a long period or decay gradually. MPs build up significantly in the soil, potentially endangering terrestrial ecosystems.

7.2.1. Effects of MPs on the soil biophysical environment

The soil ecosystem is negatively impacted by MPs in a variety of ways depending on their type, size, and shape. Wearing clothing has also been

shown to cause textiles to emit MPs into the air, which are then deposited in soil.[105] When textile waste is dumped as litter or thrown away in landfills, it can disintegrate and cause fibers to leak into the ground. Examples include single-use face masks, ropes, tarpaulins, and misplaced clothing.[106] It is believed that organisms like earthworms have the ability to move considerable quantities of MPs from the soil surface to greater depths.[105] MPs, which resemble soil particles in size and form, have less of an effect on the soil's structure.[107] Since fibrous MPs have a very different form from soil non-linear particles, it is possible that they can coil around soil particles more effectively. MPs shapes have the power to change the biophysical characteristics of the soil. According to a study, MPs can alter the soil's structure and impair its bulk density and water-holding capacity.[108] Types of MPs have an impact on how MPs and soil pore space interact. As opposed to PES beads, polyacrylic beads and PE fragments did not significantly lessen soil bulk density.

7.2.2. Effects of MPs on soil organisms

MPs pollution changes the soil's physicochemical characteristics, which impacts soil hydrodynamics and microbiological activity.[109] Different soil microorganisms are affected by MPs in different ways. The microbial populations in the soil may unavoidably change as a result, having effects on the soil's flora and fauna. MPs affect the composition of the microbial community in the soil. It was discovered that PE considerably altered the composition of bacterial communities and had an impact on their succession.[110] By altering the soil environment, MPs have an impact on the evolution of soil microorganism.[111] MPs can provide nutrients and organic carbon to the soil, and bacteria that make use of this resource have a higher chance of surviving.

Moreover, MPs can be consumed by soil animals. Consumption of MPs may cause soil animals to receive insufficient nutrients, experience oxidative stress, and suffer intestine injury, which will impede their ability to develop and reproduce.[112] Lower concentrations of MPs had an impact on the mortality of soil animals in spite of the presence of tissue and immune system damage.[8] Future studies are required to determine whether the immune system and histopathology can assist as possible indications of sensitivity to ingested MPs. The effects of MPs on soil creatures might vary, depending on their size and concentration. Through experimentation, it was discovered that the presence of MPs could have a considerable impact on the diversity and abundance of nematode and microarthropod communities.

The soil food chain may be affected by the way soil fauna reacts to the addition of MPs. The presence of MPs can drastically alter the soil

fauna through extractable chemicals in MPs.[113] MPs from the atmosphere may contaminate soil and have an impact on plant growth as well as on the leaves of the plants through dry and wet deposition. MPs present in the air can cling to leaf surfaces and even enter plants through stomata. Because of the morphology of the leaf stomata, the majority of the MPs that attached to and entered the leaf were found to be fibrous.[114] Stomata in plant leaves take in NPs and send them downward to the roots. MPs foliar absorption may impact a plant's ability to photosynthesis.[115] Foliar MPs exposure can significantly slow growth by reducing the area of leaf, weight of dry crop and height of crop.[116]

7.3. Effects of MPs on the biogeochemical cycles of soil carbon and nitrogen

The carbon and nitrogen cycles, soil microorganisms, climate change, and greenhouse gases (GHGs) are all interconnected in significant ways. According to certain research, MPs deposited into the soil can change the makeup and operation of microbial communities, which in turn can influence the carbon and nitrogen cycles in the soil.[117] According to a study, MPs additions can considerably increase CO_2 emissions from soil.[58] This might accelerate global warming and lead to additional severe environmental issues. The addition of MPs had no impact on soil nitrogen emissions, though. The size of MPs has an impact on the type of dissolved organic carbon in soil. According to a study, PVC hindered sediment nitrification and denitrification, while PUF and PLA encouraged it. The soil's ability to hold nitrogen was diminished in MPs-treated soils, as MPs modify the porosity of the soil, which could further impact the nitrogen cycle.[117] MPs could affect the carbon, nitrogen, and phosphorus-related genes and enzymes in soil bacteria, which in turn could affect the soil nutrient cycle.[118]

7.4. Effect on human health

As per studies, humans are primarily in continuous contact with environmental MPs through inhalation, ingestion, and skin contact.

7.4.1. Inhalation by respiration

Gravity causes MPs to settle in different parts of the human respiratory system. The body's concentration of MPs can be decreased through the clearance mechanism. The clearance system may be slowed down by MPs particle toxicity brought on by cytotoxicity, excessive dust load, oxidative stress, and particulate matter translocation, thus impacting the overall human health.[119] Because only MPs having sizes less than 10 μm can be

ingested, an MPs size influences its respiratory potential. Since, MPs smaller than 10 µm were found both indoors and outdoors, they can also enter the human body though inhalation. Even though mucociliary clearance occurs for the majority of the coarse inhalable particles in the upper airway, some tiny particles can evade this system and settle deep inside the lungs.[120] As, fragments are mostly transmitted by air, 31 synthetic polymer particles and fibers, of which 87.5% fragmented and 12.5% were fibers, were found in the lung tissue of 13 of the 20 autopsies of the deceased, validating this theory. The upper respiratory tract retains the majority of the MPs that are inhaled. The duration and intensity of MPs exposure also have an impact on the impacts on human health. High levels of MPs in the environment could pose serious threats to people's health.[121] A mannequin could simulate inhaling 272 MPs within 24 h using a breathing thermal manikin.[120] When it comes to exposure to airborne MPs, young children are typically more at risk than adults.[122] Low concentrations of atmospheric MPs typically expose the general public. However, practitioners who are exposed to high levels of MPs on the job, run the risk of developing occupational illnesses.[123] The size and form of MPs have a significant impact on cytotoxicity.. Lung cells exposed to MPs experienced significant morphological alterations and a reduction in cell growth.[124] Human lung epithelial cells' internalization, cell viability, cell cycle, and apoptosis may be impacted by NPs.[125] It was shown that exposure to MPs caused inflammation, oxidative damage, and disruption of intercellular junctional proteins in normal human lung epithelial cells BEAS-2B.[126] Respiratory illnesses, both acute and chronic, may develop from this. Among other things, metals, possible PAH transporters, plastic additives, and airborne MPs were suspected of producing chemical toxicity.[75]

The inflammatory potential of MPs with additives was often higher than that of MPs without additives.[123] Ultrafine particles (UFPs) may be inhaled and diffused from the lungs into the bloodstream, or they may be ingested after airway mucociliary clearance. These inhaled or ingested particles may continue to travel in the bloodstream and become harmful to other organs after translocation, including the liver, embryo, brain, etc.[121]

7.4.2. Oral ingestion

MPs are typically found in the food and water consumed by humans. Additionally, a lot of MPs are present in commonly used tap water, bottled drinks, and alcohol. In tests of tap water and various brands of bottled water from around the world, MPs particles were found.[73] 95% of MPs particles having size in the range of 6–100 µm, or 80% of MP particles with a size of 5–20 µm, were found in bottled water. Moreover, alcohol, salt, and shellfish all contain human-caused MPs pollution. Some MPs that reach the

human digestive tract are expelled through feces[127] and some are still seen in human babies. The human body can develop abrasions, perforations, and even obstructions in the digestive tract as a result of MPs ingestion.[128] All plastic particles exhibited high resistance to artificial digestive juices, as demonstrated by the effects of artificial *in vitro* digestion on five types of MPs, and the primary stages of the human gastrointestinal tract do not break down particles.[129] PE could decrease cell functions and alter the composition of intestinal flora, which would be dangerous for human health, according to *in vitro* simulation studies on human cells Caco-2 and intestinal flora.[130] On the other hand, when PS of various sizes was utilized on Caco-2 cells, the particles did not harm or inflame the cells.[131] MPs at the ambient level are unlikely to provide a severe health concern to humans when consumed orally. Epidemiological studies, *in vivo* animal experiments, and *in vitro* cell, culture techniques are used in the majority of current investigations on MPs in the human digestive system. MPs present in the human body can accrue in the digestive system or go to other organs e.g., liver, kidney, by oral consumption.[132] Since MPs get decolorized in the digestive tract, the majority of MPs samples found in human digestion were clear laments.[133] It is thus far uncertain whether the MPs pigments stick around in people's bodies and harm them.

7.4.3. Skin contact and dermal absorption

Pollutants in the form of particles have been linked to skin ageing. Particles from traffic can worsen wrinkles in the nasolabial creases and freckles on the cheeks and forehead. Nevertheless, tiny particulate matter is frequently viewed as a significant environmental hazard that leads to skin illnesses. Skin illnesses can develop and worsen as a result of various air pollution particulates that cause oxidative stress by releasing pro-inflammatory cytokines and reactive oxygen species (ROS). Further study in this area is required because there is no direct indication that atmospheric MPs can enter the skin.

8. Identification methods in MPs analysis

Monitoring of MPs in a range of biotic and abiotic environmental matrices can give fundamental scientific data on their pollution status, concentration hotspots, historical trends, and the fate and exposure of organisms. MPs monitoring studies demand accurate and comparable sampling and analytical techniques. In addition, proper qualitative and/or quantitative analytical procedures for MPs are needed for laboratory accumulation, toxicity, and weathering research. However, because MPs research is a recently reemerging topic, analytical methodologies for MPs are still in the early stages of development. MPs in environmental samples can be analyzed

using extraction, isolation (or separation), identification, and quantification techniques (or classification).[134] Traditionally, preproduction resin pellets and other large MPs in the 1–5 mm size range have been identified by visual sorting with the naked eye.[135] However, in order to successfully perform an ecological risk assessment, quantification at the smallest scale is essential due to the ubiquitous environmental presence of MPs and their harmful biological consequences.

Researchers have been unable to create a uniform classification of MPs data due to the vast range of sizes of MPs and the complexity of their colors, shapes, and polymer kinds, which makes it more challenging to compare data. Using a single analytical technique, it is challenging to completely and accurately identify MPs of different sizes, shapes, and polymer types from complex environmental matrices. Therefore, it has become common practise to combine more than two analytical methodologies. MPs examination often involves two steps: physical evaluation of putative plastics (using a microscope, for example) and chemical characterization (using a spectrometer, for example) to confirm the presence of plastics. Each technique, as well as different combinations of them, has benefits and drawbacks.

8.1. Visual methods

Large MPs (1–5 mm) were commonly found on beaches and, to a lesser extent, in surface waters before the term "MPs" gained popularity. The size range of MPs is rather wide, hence sorting and identification was typically done side by side in a tray using forceps and the human eye. This visual method can be used to identify some small plastics, and colorful plastic pieces and pre-production resin pellets that vary in size from 2 to 5 mm and can be recognized by sight. There is a greater likelihood that small plastic particles will be missed by sorting in high strandline samples taken from beaches with high concentrations of interfering inorganic and organic materials, or in ambiguous plastics (approximately 1 mm in size) with colors similar to the interfering particles. However, for both specialists and nonprofessional volunteers who have undergone brief instruction, visual sorting and identification of big MPs offers an easy, straightforward, and quick procedure.[136]

8.2. Stereo (or dissecting) microscopy

A popular method for identifying MPs with sizes in the hundreds of micron range is stereo- (or dissecting) microscopy (e.g., neuston net samples).[137] Microscopy-enhanced photos give objects' surface textures and

structural details in great detail, which is crucial for recognizing confusing, plastic-like particles. Although most particles in this size range can typically be identified by microscopy, it can be challenging to confidently identify plastics in sub-micron-sized particles (<100 µm) that lack a characteristic color or shape.[138] The microscopic detection of MPs on filter paper can be hampered by sediment samples whose light sediment particles are poorly separated by density. Microscopic inspection is further complicated by biogenic elements from silt and neuston net samples that have not been entirely removed by chemical digestion. According to numerous field monitoring studies, the majority of the MPs in water, sediment, and biota were made up of fibres.[139] However, by using only a microscope, it was challenging to distinguish between fibers that were synthetic (like polyester) and natural (like colored cotton).

8.3. Scanning electron microscopy (SEM)

Images of plastic-like particles can be seen at extremely high magnification and clarity using scanning electron microscopy (SEM). The separation of MPs from organic particles is made easier by high-resolution pictures of the surface roughness of the particles. The elemental makeup of the same object is revealed by additional investigation using energy-dispersive X-ray spectroscopy (EDS). Carbon-dominant plastics can be distinguished from inorganic particles by looking at the elemental makeup of the particles. The number of samples that can be processed using SEM-EDS is constrained by its cost and the time and effort required for sample preparation and analysis. Plastic colors cannot be utilized in SEM as identifiers. The technique is advised for additional elemental analysis and surface characterization of particular plastic particles.[139]

In some instances, other cutting-edge microscopy methods have been utilized to identify plastic particles. In laboratory accumulation and toxicity investigations, polyethylene (PE) particles were successfully identified using polarized optical microscopy. Polarized light transmission can be affected by the crystal structure of plastics, and this transmission can then be measured.

8.4. Fourier transform infrared (FTIR) spectroscopy

A FTIR can be used for the identification to disclose the polymer composition as well as the abundance of MPs, which can assst in identifying identifying the source. The weathering status of MPs is also shown by the proportional content of oxygenated bonds (such as carbonyl groups) in the IR spectra. By using this technology, it is simple to identify

carbon-based polymers, and the distinctive spectra produced by various bond compositions, which allow the plastics to be distinguished from other organic and inorganic particles. A reliable polymer spectrum library permits the identification of particular polymer types in addition to the confirmation of plastics. Micro-FTIR (m-FTIR) is used to conduct microscopic observation of micro-sized plastic-like particles before spectroscopic identification on a single platform by shifting between the object lens and IR probe, which is necessary for small MPs. FTIR measurement for MPs is accessible in transmission, reflectance, and attenuated total reflectance (ATR) modes.[140] The sample preparation step is not necessary for thick and opaque MPs in the reflectance and ATR modes, in contrast to the transmission mode. Additionally, the ATR mode creates stable spectra from uneven MPs surfaces, whereas the reflectance mode creates unstable spectra. Theoretically, it is possible to detect MPs as small as the ATR probe's IR beam aperture (10 μm, for example). With the identification of plastic-like particles by microscopy and subsequent chemical confirmation by spectroscopy, m-ATR-FTIR now offers a promising way to identify MPs in environmental samples. However, it can be challenging to get distinct spectra for MPs with a maximum length of 50 μm, necessitating numerous attempts. Surface contact analysis is a type of measurement used in ATR-FTIR. MPs that are extremely worn or delicate may be damaged by the pressure from the ATR probe, and small plastic particles may be attracted to or pushed off the filter paper by electrostatic interaction with the probe tip. When used in contact analysis, remains of hard, sharp inorganic particles from sand samples can easily harm an ATR probe manufactured of germanium. By employing the IR spectra to identify each plastic-like particle, it is possible to limit the possibility of missing plastic particles and false-positive quantification of non-plastic particles. m-FTIR equipment is quite pricey. It takes a lot of time to use an ATR probe to identify each and every plastic-like particle. Samples may contain a large number of moderately or severely weathered MPs that are often formed of complicated and occasionally composite materials. Therefore, to acquire clear spectra and correctly interpret them, an expert operator is needed. A professional judgment based on a detailed examination of the spectrum is required when weathered plastics exhibit spectra of different polymers that have poor hitting potential in libraries. This may result in decisions that are made arbitrarily or based on statistical data and the criteria for meeting the percentage values. The human labor required in the FTIR process can be reduced by using pre-programmed, semi-automatic mapping without the requirement for microscopic pre-selection of particles for analysis.

It is uncommon to employ single FTIR mapping for routine examination of complicated field samples since it requires longer operating durations

and only covers a small region for the chosen wavenumbers of certain polymer types. When compared to single beam mapping, the shortcomings of the focal plane array (FPA)-based reflectance imaging method could be approved. This method could provide information for the identification of MPs (150-250 µm) on larger surface areas, more quickly, and without sacrificing spatial resolution. Although FTIR operation time is greatly reduced by semi-automatic mapping technology, it still takes at least 9 h to scan a single filter paper. To enable frequent use of this approach in MPs analysis, other challenges must be resolved, that include the potential for plastics to be aggregated and the potential for plastics with irregular shapes to cause refractive errors in the reflectance mode. Therefore, after the analysis, each location and spectrum must now be manually validated. Smaller plastics (e.g., <100 µm) need additional identification due to the reflectance mode's limits in producing clear spectra for tiny plastic particles. This needs more time for identification by ATR-FTIR or microscopy.

8.5. Raman spectroscopy

Micro-plastics have also been identified by using Raman spectroscopy. Each polymer has a distinct spectrum due to the molecular structure and atom composition that influence the backscattered light frequencies produced when a laser beam strikes an item. Raman analysis not only distinguishes between different types of plastics but also offers profiles of each sample's polymer makeup, much like FTIR. Raman spectroscopy is analogous to the FTIR method when it comes to the nondestructive chemical analysis and microscopy combination, including the need for pricey equipment. In difficult MPs identification, the various reactions and spectra from FTIR and Raman spectroscopy can conflict with one another. Raman spectroscopy, which uses a narrower laser beam than FTIR, can identify MPs as small as a few millimeters in size. The advantage of Raman spectroscopy's non-contact examination is that the MPs samples are preserved for potential future analysis. MPs in zooplankton samples can be found using confocal microscopy and Raman spectroscopy. Raman spectroscopy is sensitive to the addition and pigment compounds included in MPs, which makes it difficult to distinguish between different types of polymers.

8.6. SERS (surface-enhanced Raman spectroscopy)

SERS is a rapid, noninvasive spectroscopic analytical technique that uses distinct molecular vibrational fingerprints to identify analytes at the single-molecule level.[141] Most molecules in a typical Raman scattering process produce weak Raman signals due to their relatively lower Raman scattering

cross-section. Increasing the Raman signals to greater than or equal to the fifth order in SERS requires the use of plasmonic nanostructures composed of Ag, Au, Cu, and Al. The primary mechanism behind this amplification is the activation of localized surface plasmons, which came about as a result of incoming visible light interacting with noble metallic nanoparticles.[142] Specifically, light that corresponds with the electron oscillation frequency produces a localized surface plasmon resonance, which increases the electrical field at the particle surface.[143] Instead of being uniformly distributed across the plasmonic nanostructure, the produced electromagnetic field is typically limited to sharp edges and tiny interparticle gaps, creating plasmonic "hotspots." Moreover, the elevated electromagnetic field intensifies both excitation and emitted radiation, leading to the notable increase in SERS signals that are independent of analyte type. This phenomenon, known as electromagnetic SERS enhancement, is one of the two established theories for SERS signal amplification.[141] One of the challenges in SERS NPs detection might be optimizing the plasmonic hotspot for a strong SERS signal. It is commonly known that structures with sharp edges, are more effective at enhancing scattering strength. Therefore, employing optimally sized and shaped plasmonic nanoparticles to fabricate SERS substrates could be a great option.[144,145]

When utilizing SERS, it is essential to choose an appropriate laser power level. It is well known that, depending on the intensity of the laser, the temperature surrounding a hot spot can rise to much greater temperatures than 200 °C. This temperature is greater than most common polymers' glass transition temperature. To overcome these challenges, it is required to optimize the procedure by regulating the laser's power, analyte types, and excitation wavelength.[144]

8.7. AFM-IR

AFM-IR, or AFM with infrared spectroscopy, is a relatively new method in the sciences that examines events at the nanoscale. In AFM-IR, an AFM scans the surface of the material to provide an image with a high spatial resolution of as low as 50 nm, whereas infrared spectroscopy elucidates the mechanical and thermal properties of MPs. For the detection of PE, PET, and polycaprolactone with a particle size of less than 500 µm, this approach has shown noteworthy success. With spatial resolution below the traditional optical diffraction limits, it is utilized for compositional mapping and chemical analysis. With remarkable precision and detail, this approach provides physiochemical characteristics. It enables further MPs and NPs characterization, including hydrophobicity and conductivity. The

technique requires a sample pre-processing step to make presumed MPs accessible to AFM.

8.8. Mass spectroscopy

Using mass spectrometry in conjunction with gas and liquid chromatography, one can analyze MPs by examining the distinct products that are produced during the pyrolyzed or hydrolyzed of MPs. These mass approaches have considerable potential for MPs identification and quantification, especially NPs, due to their outstanding sensitivity.

8.8.1. Inductively coupled plasma mass spectroscopy

The development of isotope tagging technology resulted in the invention of single particle inductively coupled plasma mass spectroscopy (ICP-MS), a novel mass spectrometry method for measuring MPs. This innovative method has made it feasible to measure the size and quantity of nanoparticles. For metal-based nanoparticles in environmental materials, this technique is often employed.[146] For instance, at a comparatively low detection limit of 8.4×10^5 particles/L, ICP-MS has been used to evaluate the size and number concentration of the model Au-coated MPs (at submicrometer scale). Nevertheless, its utilization necessitated extensive sample preparation due to its reliance on an indirect evaluation of the Au coating.[147] It has been shown that ICP-MS can assess model MPs particle sizes and number concentrations by tracking 13 C during the process.[148]

8.8.2. Pyrolysis gas chromatography–mass spectrometry (Pyr-GC–MS)

Mass spectrometry-based destructive analysis following the pyrolysis of suspected polymers is becoming more and more popular. This technique employs the thermal degradation of MPs in an inert atmosphere to reveal the structural composition of a macromolecule. Sample preparation for this approach is minimal, and direct injection is a possibility based on the environmental matrix. This method overcomes the problem of potential plastic underestimation because it is employed regardless of particle size and provides information on the mass of polymers in the samples.[149]

When analyzing complex matrices, pretreatment procedures that include digestion or separation before injecting the material in the Pyr-GC-MS may be error-prone. As a result, the combination of pressurized liquid extraction (PLE) and other techniques has emerged as a potentially effective method to eradicate certain polymers (PVC, PMMA, PS, PE and PP) from intricate environmental samples. Because it can be difficult to prevent interference when working with environmental samples, double-shot pyrolysis in conjunction with PLE has also been employed.[150] In addition to polymer types,

Pyr-GC-MS may identify plastic additives in samples from the first shot, when low-temperature thermally desorbs the interference. But the second shot contains the real polymer identification and quantification. It can provide an estimate of the mass of polymers in each sample rather than a particle count. Pyr-GC-MS, like other methods without optical interrogation, does not reveal the physical form of plastic waste.[151]

8.8.3. Liquid chromatography–tandem mass spectrometry (LC–MS/MS)

Minimal quantities of polymers can be detected using LC-MS/MS. This a versatile instrument used in medicine, pharmacology and the assessment of environmental exposures to minute heat-labile, nonvolatile chemicals. Polycarbonate, PET and nylon-based MPs were successfully identified and quantitatively analyzed in sewage sludge using LC-MS/MS.[152,153] This method can be detrimental since it necessitates the depolymerization of macromolecules before examination. The mass and quantity of monomers produced during depolymerization are disclosed by the detection method; however, the size, number, color, and shape of the MPs remain unknown. The method can also yield information about the size profile distribution when combined with size fractionation. Additionally, this method necessitates a substantial financial investment, knowledgeable operators, and meticulous sample preparation—which is usually vulnerable to experimental contamination.[154]

8.8.4. Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry

MALDI-TOF is a mass spectrometry technique that uses an ion source of MALDI and a detector of TOF as the mass analyzer. MPs investigations may examine this technique, which has been utilized since 1996 to detect whole bacterial cells without pretreatment. The device has proven to be effective in differentiating between different types of polymers. In conjunction with imaging techniques, it can also make it easier to examine the shape, size, color, and composition of MPs. More specifically, MALDI-TOF is a crucial tool for complex polymer characterization. Recently, PS and PET in complex biological materials have been identified and quantified using this mass spectrometric approach.[155] Perhaps one of the preferred analytical techniques is this one due to its wide mass range of detection.[156]

8.9. Shrinking surface bubble deposition (SSBD)

The Shrinking Surface Bubble Deposition (SSBD) method is an innovative way to identify MPs in an array of environmental samples. The method concentrates the suspended particles in the water and deposits them on a substrate for observation and characterization by using plasmonic heating

of mixed silver NPs with ambient water samples to produce a bubble through laser excitation. This method enables the direct visualization and identification of NPs in a range of water samples.[157]

8.10. Flow cytometry

Flow cytometry is an extremely sensitive tool for detecting fluorescence signals and light scattering, which enables quick quantification of fluorescence-stained (NR, for example) MPs. It has been widely applied to the detection and analysis of MPs, such as molecules, cells, and microbes in liquid samples that are hydrodynamically concentrated into a stream so that they flow past incident light beams from one or more lasers one at a time. Sensitive photomultiplier tubes monitor parameters including light scatter and fluorescence, providing information about surface roughness/granularity (side scatter), size (forward scatter) and fluorescence from pigments or dye markers. A significant benefit of flow cytometry, aside from its sensitivity, is the elimination of human subjectivity because the method does not necessitate the visual interpretation of particles.[158,159]

8.11. Thermal analysis

Recently, the thermo-analytical method—which gauges changes in polymer physical and chemical characteristics based on their thermal stability—was put to the test for identifying MPs. The technique of differential scanning calorimetry (DSC) can be used to examine the thermal characteristics of polymeric materials. Each plastic product has unique DSC characteristics; hence the approach needs reference materials to distinguish between different polymer kinds. DSC can be helpful in locating particular principal MPs for which reference materials are accessible, like polyethylene microbeads. Polyvinyl chloride (PVC), polyamide (PA), polyester (PES), polyethylene terephthalate (PET), and polyurethane (PU) could not be detected by thermogravimetry (TGA) in combination with DSC because their phase transition signals overlaps. TGA combined the benefits of a larger sample size with TGA compared to pyrolysis (pyro)-GC-MS and higher resolution with GC-MS compared to DSC when paired with solid-phase extraction (SPE), which was then coupled with thermal desorption gas chromatography mass spectrometry (TDS-GC-MS). The TGA-SPE-TDS-GC-MS approach was used to quantitatively identify PE injected into soil and mussel samples. A polypropylene (PP), polystyrene (PS), and polymer combination were also examined using this method.[139]

Thermal analysis, however, is a destructive technique that precludes the further study of the MP sample. Although DSC analysis is quick and

relatively easy, it is not always successful in distinguishing MPs from other polymer products in environmental samples. Chemical additives in MPs can be simultaneously analyzed using thermal analysis and GC-MS. Bulk samples, and single particles, can be examined using GC-MS and thermal analysis, which gives weighted (w/w) concentration data for MPs. The bulk analysis does not, however, offer data on the quantity, size, and shape of the MPs that were analyzed. Contrary to FTIR and Raman spectroscopy, this method demands a skilled operator as well as significantly more time and effort for operating the instrument and data analysis. These techniques would be helpful for screening analyses of bulk samples or further corresponding investigations of MPs that have not yet been thoroughly characterized by spectroscopy, but it seems too soon to employ them for routine analysis of MPs from environmental samples.

8.12. Material-supported MPs/NPs detection

8.12.1. Nanopipettes

MPs have grown in importance during the previous 20 years, and numerous research have examined MPs/NPs in environmental sample data. Scientists have conducted extensive research on this subject and have developed a range of innovative analytical techniques for their detection.[160–162] It has recently been reported that the analytical community has several potential uses for nanopipettes, which are made by heating glass capillaries during the pull-out process. Nanopipettes became popular as a means of developing chemical and biological sensors and conducting ion transport investigations at the nanoscale.[163,164]

8.12.2. Microfluidic chips

To swiftly identify MPs in environmental samples, a unique analytical method still needs to be developed and optimised.[165–167] Micro-reservoirs leading to microfilters that are intended to capture all MPs particles in an ultra-compact region were used in the fabrication of the micro-optofluidic platform. To identify the type and nature of MPs present in this ultra-compact space, the researchers were able to successfully enable optical spectroscopy and imaging. Furthermore, passive size sorting was used to separate the MPs based on the size ranges in various reservoirs. The distributed sizes of the MPs samples were recovered using a reference technique that included flow cytometry; however, information regarding their chemical nature was lost in the process. In order to verify the possible use of this micro-optofluidic platform, model samples comprising standard MPs with varying sizes and chemical compositions were combined and effectively examined.[168]

9. Global initiatives and control measures by international organisations, regional unions, and associations

Globally, there is increasing concern about regulating MPs and NPs, and several nations have responded to this concern by enacting different laws and regulations. An outline of a few of these policies is provided below (Table 4). In an effort to lessen the negative effects that MPs and NPs have on the environment and human health, these policies combine voluntary actions, regulatory strategies, and international cooperation. The efficacy of these interventions varies and frequently necessitates continual evaluation and accounting in response to novel scientific findings and technological advances.

10. Recommendations for future research

The pollution of terrestrial and marine ecosystems by MPs is widespread worldwide. Research on a number of significant sources and release channels for MPs has been compiled in this review. Significant MPs loadings are anticipated across a wide range of environmental situations, which suggests a larger yearly discharge than the figure for the marine environment. After intensive literature survey it was found that there are various gaps in management of MPs which present numerous opportunities for future research. The distribution, movement, and destiny of MPs particles are predicted to be influenced by a variety of dynamic processes in terrestrial and freshwater systems because of their enormous complexity. A series of precise guidelines for further research have been identified by this review, as:

10.1. Harmonic progression of approaches and reporting, with improved quality assurance and control (QA/QC) techniques, to assure adequate data quality and enable accurateness between datasets

The vast range of methodological approaches used in the review for the investigations of the presence of MPs in terrestrial and marine ecosystems, including samples of biota, sediment and water is notable. This includes differences in the methods used for sampling, treating samples, using analytical procedures, and analyzing different particle size classes. Additionally, it is impossible to evaluate the quality of given data because many studies do not use a comparable set of QA/QC methods. Findings are frequently presented in a variety of ways, such as by using various measurement units or by publishing simply summary statistics that vary as well (such as minimum/maximum, mean, or median). The inability of diverse research to be compared is the pinnacle of this variability. In

Table 4. Different policy/strategies/initiatives adopted by the different national/international organisations, regional unions and associations.

National/ International	Policy	Goals/ Initiatives	Measures	Ref.
European Union (EU)	EU Plastics Strategy	Make all plastic packaging recyclable by 2030, reduce plastic waste, promote circular economy practices.	Bans on certain single-use plastics, mandatory recycled content in products, extended producer responsibility schemes, and incentives for innovation in plastic recycling technologies.	[169]
	REACH Regulation	Regulates chemicals and their safe use, including those that might contribute to MPs pollution.	Requires registration, evaluation, and restriction of substances, including additives in plastics that can degrade into MPs.	[170]
	(MSFD) Marine Strategy Framework Directive	EU maritime waterways must reach Good Environmental Status (GES) by 2020.	Member states must develop marine strategies including measures to reduce marine litter, monitor MPs levels, and assess the effectiveness of policies.	[171]
United States	Microbead-Free Waters Act of 2015	The production and distribution of rinse-off cosmetics with plastic microbeads are forbidden under federal legislation.		[172]
	NOAA Marine Debris Program	Addresses marine debris through prevention, removal, research, and outreach.	Funding for research on MPs, educational campaigns, and partnerships with local organizations for cleanup efforts.	[173]
Canada	Microbeads in Toiletries Regulations	Plastic microbeads in toiletries are prohibited nationwide from being developed, imported, or sold as of July 1, 2018.		[174]
	Canada-wide Strategy on Zero Plastic Waste	Eliminate plastic waste and reduce plastic pollution	Enhancing extended producer responsibility, improving waste management systems, and supporting innovation in plastic alternatives.	[175]
Japan	Act on Promotion of Resource Circulation for Plastics:	Comprehensive approach to plastic waste, promoting recycling and the use of biodegradable plastics	Encourages businesses to develop and use alternatives to traditional plastics and improve recycling infrastructure	[176]
	Law for the Promotion of Marine Litter Processing	Targets reduction of marine litter including MPs	Provides funding for local governments to clean up marine litter and improve waste management practices	[177]

(*Continued*)

Table 4. Continued.

National/ International	Policy	Goals/ Initiatives	Measures	Ref.
Australia	National Waste Policy	Reduce waste and improve recycling by 2030	Phasing out problematic and unnecessary plastics by 2025, promoting product stewardship, and enhancing waste management infrastructure.	[178]
China	Plastic Waste Management Policies	Limitations on microbeads in cosmetics and nationwide prohibitions on single-use plastics	Phased approach to banning and reducing single-use plastics, improving recycling rates, and promoting biodegradable alternatives	[179]
	Action Plan for Zero-Waste Cities	Pilot initiative aimed at creating sustainable waste management practices	Focus on reducing plastic pollution, enhancing waste segregation and recycling, and promoting public awareness	[180]
South Korea	Regulations on Microbeads	Ban on the manufacture, import, and sale of personal care products containing plastic microbeads, effective July 2017	–	[181]
	Marine Litter Management Plan	National strategy to reduce marine litter, including MPs	Regular cleanup activities, public awareness campaigns, and partnerships with fishing and shipping industries to reduce litter	[182]
International Initiatives	United Nations Environment Programme (UNEP)	Global initiatives to combat plastic pollution	The Clean Seas campaign, promoting policy changes, and supporting national and local initiatives to reduce plastic waste	[183]
	International Convention for the Prevention of Pollution from Ships (MARPOL)	Includes regulations to prevent pollution from ships, addressing both operational and accidental discharges of plastics	Amendments to MARPOL Annex V prohibit the discharge of all plastics from ships	[184]
	Global Partnership on Marine Litter (GPML)	A multi-stakeholder partnership to tackle marine litter	Facilitates international collaboration, shares best practices, and supports research and policy development to reduce marine litter, including MPs	[185]

addition to providing useful baselines that can monitor the influence of remediation or reduction measures, collaboration of analytical techniques and investigating formats and the publication of data in appropriate repositories will aid in reducing ambiguity in a comprehensive, global overview of the condition of contamination.

10.2. An exhaustive analysis of the fate, sources, and effects of MPs in agricultural environments

The confluence of numerous MPs particle sources and release channels can be seen in agricultural contexts. Although the extent and relevance of this are mostly unclear, particles may also be linked to greater chemical loads from plastic additives (mostly for reducing the photodegradation of mulching films) or sorbed contaminants (such as from wastewater treatment plants). It is challenging to evaluate the relative contributions of various sources under pertinent environmental conditions or the accumulation of particles over time because agricultural settings are relatively uncontaminated and the destiny of particles in agricultural soils is still poorly understood. The effects of MPs contamination in agricultural ecosystems have only recently been the subject of a tiny amount of research. To get a comprehensive understanding of the dangers presented by MPs over spatial and temporal domains, more research is necessary. This is especially crucial considering the possibility that any adverse consequences could affect the security of the food supply and the health of the land.

10.3. Assessment of road-associated MPs particles (RAMP) as a source of MP to the environment

There are significant awareness gaps regarding environmental loadings, transportation from the road to various matrices, and persistence in gully-pots and water treatment systems for MPs particles associated with the road. In order to precisely estimate the amount of these road-associated microparticles released into the environment, including the contributions from various road-related sources, more investigation is urgently required. This is relevant in view of the recent attention received by these road-associated microparticles in a number of analyses of the worldwide major sources of MPs in the environment. Through the use of a novel and improved method for analyzing these microparticles in environmental samples, quantification should be possible.

10.4. Measures and technologies to separate particles within waste water treatment plant system or to reduce MPs emissions to wastewater

The role of wastewater systems, including their plants and combined sewer overflows, as a release channel for MP particles into both terrestrial and marine environments, was underlined in this research. In order to lessen the load on these plants and restrict releases from untreated discharges like combined sewer overflows, attempts should be made to minimize the amount of MPs particles entering wastewater at the source. A large portion

of the wastewater produced worldwide is released untreated and is not connected to a WWTP. The environmental emission of MPs in so many nations would be constrained by improvements in wastewater treatment capacity worldwide. Many MPs forms have sewage sludge application on land recognized as a key release mechanism.

10.5. A Better perspective of the controls sustaining the preservation and transport of MPs particles in aquatic systems, including more accurate flux estimates to the marine environment

From their release to their final destination, MPs particles in freshwater systems are expected to travel over a complicated course. Several processes that obstruct downstream transit may be included in this. To establish thresholds and controls on MPs transport in freshwater environments, these dynamics need to be studied further. The bulk of particles is predicted to eventually end up in the marine environment, via fluvial systems. In order to incorporate the necessary complexity and pinpoint the effective constraints on MP's release to the marine environment, estimates for this flow must be based on process-based research. As well as identifying zones of MPs build-up in freshwater systems, further research on the dynamics of MPs movement and spatial patterns of contamination can also highlight the places most at risk from potential contamination-related negative effects. This will aid in concentrating efforts to safeguard freshwater environments.

10.6. Further study of the occurrence of MPs in terrestrial and aquatic organisms, with an explicit focus on particle types such as RAMP

The impact of MPs pollution on ecosystems and human health has prompted extensive ecotoxicological studies. However, there is a lack of evidence regarding the uptake of MPs in actual field settings. To address this, a more comprehensive analysis of particle absorption in actual environmental conditions is essential to provide a contextual understanding of ecotoxicological investigations and to enhance risk assessments. Exposure to MPs is a critical factor in evaluating risk, and their presence in the environment can have detrimental consequences for both ecosystems and human health. Furthermore, existing studies have primarily focused on the impact of individual materials, with limited consideration for mixtures of biodegradable MPs and MPs at concentration ratios that reflect real environmental circumstances. It is imperative for future studies to incorporate such mixtures to better understand the combined effects and interactions of different types of MPs. Moreover, the concentration of MPs used in many experiments exceeds those found in the actual environment.

Therefore, the concept of "environmentally relevant concentration" has become a significant consideration in toxicity experiments. Additionally, the duration of exposure is a crucial factor, as the continuous exposure to MPs in natural environmental conditions and the slower depuration process compared to other particles can significantly influence the outcomes of ecotoxicological studies.

10.7. Assessment of environment sinks for MPs particles at spatial and temporal scales

This analysis identified a number of potential locations that, in terrestrial and freshwater settings, could serve as transient, long-term, or even permanent sinks for MPs particles. Initial research has looked at the accumulation of some of these particles and their possible residence times. However, study is needed to determine the spatial and temporal scales on which these environments act as stores, as well as the possibility that they could later become sources of MPs contamination through reworking and remobilization. This is required to develop more appropriate and better-targeted remediation procedures rather than relying on quick solutions and acquire a better long-term perspective on environmental damage.

11. Comprehensive strategies for mitigating MPs pollution in textile lifecycles

11.1. The design and production pathway

One strategy for reducing microfibre shedding has been suggested: switching textile patterns to use more natural fibers. Besides this, the increased release of microfibres may be caused by the manufacturing procedures used to create synthetic fibers, yarns, textiles, and goods. Particularly, an essential component in the development of MPs is the use of abrasive friction during manufacture. Microfibre release while using a product could be decreased by employing alternate production techniques or textile construction techniques.

11.2. The use and caretaking pathway

The requirement that all washing machines have a specific microfiber filter by January 2025 was first introduced in France in 2020. Utilizing a washing machine filter can cut the emission of microfibres by as much as 80%. The release of microfibers is also impacted by the utilization of detergent and fabric softener. By designing non-aggressive, liquid detergents that are efficient at low temperatures and don't rinse off textile finishes,

some of which guard against fiber breaking, detergent producers can help reduce the shedding of microfibers. Powder detergents shouldn't be used on synthetic fabrics because they create friction and break fibers.

11.3. Change in consumer behavior

According to numerous studies, the first few washes of new clothing are when microfibre releases are particularly high. This indicates that fast fashion, in which clothes are worn briefly before being replaced frequently, is responsible for a significant portion of microfibre releases. Therefore, modifications in consumer purchasing habits are required, as well as understanding of the effect that new apparel has on MPs pollution. More regenerative business models, which encourage less consumption and longer use, could allow such a transformation. Reused items typically shed fewer microfibres when washed than new ones, which has the added benefit of reducing both the quantity of new purchases and trash generation. Moreover, preparing an old article for reuse requires fewer resources than the production of a new similar one, thus providing a further environmental benefit.

11.4. The disposal and end-of life processing pathway

In addition to promoting textile reuse and recycling, which would decrease MPs emissions during production and use, textile waste collection and end-of-life treatment can also reduce secondary MPs contamination by preventing littering, improperly managed waste, and textiles being blown by the wind from open landfills. Due to improper disposal, the dumping of trash from open landfills, or the washing of synthetic clothing in untreated waste water, there is a risk that microfibres will spread.

11.5. Increasing knowledge and awareness

To promote research aimed at filling these information gaps on the release, spread, and effects of NPs and MPs, continued public and corporate support is required. To solve the complex difficulties related to various primary MPs contamination from textiles, intensive interdisciplinary collaboration across technical, behavioral, and regulatory measures is required.

12. Conclusions

Concern over MPs in the environment has recently come under fire for being a distraction from other problems endangering the health of the

global environment. The impact of plastics on the environment is undoubtedly complex, and innovative approaches are needed to find answers. The fact is that, the outlaw or replacement of plastics is nearly imposible, because of their beneficial qualities and our growing reliance on them. The concern is that both their improper management and expansion into the natural environment are growing, as is their global utilization. Over time, these plastics will break down into MPs and NPs. The effects of these on toxicology or the ecosystem are not fully understood at this time. To comprehend how organic chemical partition coefficients to plastics are changed in the presence of sediment and soil, studies on the dynamic interactions between plastic particles, plasticizer additives, and environmental pollutants also need to be increased. In order to comprehend the mechanisms governing the bioaccumulation of plasticizers and co-transported compounds, studies of chemical dynamics within the guts of organisms are also required. Further investigation into MPs is required to determine their global distribution and potential human health exposure. This requires additional field sampling and the application of standardized analytical techniques.

Acknowledgment

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Raminder Kaur http://orcid.org/0000-0001-7469-375X

References

[1] Ritchie H, Roser M. *Plastic Pollution*. Our World Data.https://ourworldindata.org/plastic-pollution. 2018.

[2] Kiran BR, Kopperi H, Venkata Mohan S. Micro/nano-plastics occurrence, identification, risk analysis and mitigation: challenges and perspectives. *Rev Environ Sci Biotechnol*. 2022;*21*(1):169–203. doi:10.1007/s11157-021-09609-6.

[3] Napper IE, Thompson RC. Plastic debris in the marine environment: history and future challenges. *Glob Challenges*. 2020;*4*(6):1900081. doi:10.1002/gch2.201900081.

[4] Bashir SM, Kimiko S, Mak C-W, Fang JK-H, Gonçalves D. Personal care and cosmetic products as a potential source of environmental contamination by microplastics in a densely populated Asian city. *Front Mar Sci*. 2021; *8*:1–11. doi:10.3389/fmars.2021.683482.

[5] Yang H, Dong H, Huang Y, Chen G, Wang J. Interactions of microplastics and main pollutants and environmental behavior in soils. *Sci Total Environ.* 2022;*821*:153511. doi:10.1016/j.scitotenv.2022.153511.
[6] Mdlalose L, Chimuka L. Mitigation approaches to prevent microplastics effects in the aquatic environment: exploration of microbeads from personal care and cosmetic products. *Int J Environ Res.* 2022;*16*(5):84. doi:10.1007/s41742-022-00464-z.
[7] Chatterjee S, Sharma S. Microplastics in our oceans and marine health. *F Actions Sci Reports.* 2019; Special Issue 19 :54–61.
[8] Li J, Song Y, Cai Y. Focus topics on microplastics in soil: analytical methods, occurrence, transport, and ecological risks. *Environ. Pollut.* 2020;*257*:113570. doi:10.1016/j.envpol.2019.113570.
[9] Dris R, Gasperi J, Mirande C, et al. A first overview of textile fibers, including microplastics, in indoor and outdoor environments. *Environ. Pollut.* 2017;*221*:453–458. doi:10.1016/j.envpol.2016.12.013.
[10] Zhang B, Yang X, Chen L, Chao J, Teng J, Wang Q. Microplastics in soils: a review of possible sources, analytical methods and ecological impacts. *J Chem Tech Biotech.* 2020;*95*(8):2052–2068. doi:10.1002/jctb.6334.
[11] Campanale C, Stock F, Massarelli C, et al. Microplastics and their possible sources: the example of Ofanto River in Southeast Italy. *Environ Pollut.* 2020;*258*:113284. doi:10.1016/j.envpol.2019.113284.
[12] Allen S, Allen D, Phoenix VR, et al. Atmospheric transport and deposition of microplastics in a remote mountain catchment. *Nat Geosci.* 2019;*12*(5):339–344. doi:10.1038/s41561-019-0335-5.
[13] Rochman CM. Microplastics research from sink to source. *Science.* 2018;*360*(6384):28–29. doi:10.1126/science.aar7734.
[14] Eerkes-Medrano D, Thompson RC, Aldridge DC. Microplastics in freshwater systems: a review of the emerging threats, identification of knowledge gaps and prioritisation of research needs. *Water Res.* 2015;*75*:63–82. doi:10.1016/j.watres.2015.02.012.
[15] Horton AA, Walton A, Spurgeon DJ, Lahive E, Svendsen C. Microplastics in freshwater and terrestrial environments: evaluating the current understanding to identify the knowledge gaps and future research priorities. *Sci Total Environ.* 2017;*586*:127–141. doi:10.1016/j.scitotenv.2017.01.190.
[16] Näkki P, Setälä O, Lehtiniemi M. Bioturbation transports secondary microplastics to deeper layers in soft marine sediments of the Northern Baltic Sea. *Mar. Pollut. Bull.* 2017;*119*(1):255–261. doi:10.1016/j.marpolbul.2017.03.065.
[17] Lassen C, Hansen SF, Magnusson K, et al. Microplastics occurrence, effects and sources of releases to the environment in Denmark. Denmark: Danish Environmental Protection Agency; 2015. Report No. 1793/2015: 1–205.
[18] Heilig ML. United States patent office. *SIGGRAPH Comput Graph.* 1994;*28*(2):131–134. doi:10.1145/178951.178972.
[19] Sundt P, Per-Erik Schultze FS. Sources of microplastic- pollution to the marine environment. Miljodirektoratet: *Mepex Nor Environ Agency*; 2014. Report No. M-321/2015: 86
[20] Essel R, Engel L, Carus M and Ahrens RH. Sources of microplastics relevant to marine protection in Germany. Germany: Federal Environment Agency; 2015. Texte 64/2015, Project No. 31969, Report No. (UBA-FB) 002146/E: 1–45.
[21] Eriksen M, Mason S, Wilson S, et al. Microplastic pollution in the surface waters of the Laurentian Great Lakes. *Mar Pollut Bull.* 2013;*77*(1–2):177–182. doi:10.1016/j.marpolbul.2013.10.007.

[22] Leslie HA. Review of microplastics in cosmetics. *IVM Inst Environ Stud*. 2014;*476*(July):33.
[23] Yurtsever M. Glitters as a source of primary microplastics: an approach to environmental responsibility and ethics. *J Agric Environ Ethics*. 2019;*32*(3):459–478. doi:10.1007/s10806-019-09785-0.
[24] Kibria MG, Masuk NI, Safayet R, Nguyen HQ, Mourshed M. Plastic waste: challenges and opportunities to mitigate pollution and effective management. *Int J Environ Res*. 2023;*17*(1):20. doi:10.1007/s41742-023-00507-z.
[25] Whitacre DM, ed. *Reviews of Environmental Contamination and Toxicology*, USA: Springer Cham, 2014; Vol. 227.
[26] Rillig MC. Microplastic in terrestrial ecosystems and the soil? *Environ Sci Technol*. 2012;*46*(12):6453–6454. doi:10.1021/es302011r.
[27] Luo Y, Awoyemi OS, Naidu R, Fang C. Detection of Microplastics and Nanoplastics Released from a Kitchen Blender Using Raman Imaging. *J. Hazard. Mater.* 2023, *453*, 131403: 1-9. doi: 10.1016/j.jhazmat.2023.131403.
[28] Wang T, Li B, Zou X, et al. Emission of primary microplastics in mainland China: invisible but not negligible. *Water Res*. 2019;*162*:214–224. doi:10.1016/j.watres.2019.06.042.
[29] van Wezel A, Caris I, Kools SAE. Release of primary microplastics from consumer products to wastewater in the Netherlands. *Environ Toxicol Chem*. 2016;*35*(7):1627–1631. doi:10.1002/etc.3316.
[30] Xia B, Sui Q, Du Y, et al. Secondary PVC microplastics are more toxic than primary PVC microplastics to *Oryzias melastigma* embryos. *J Hazard Mater*. 2022;*424*:127421. doi:10.1016/j.jhazmat.2021.127421.
[31] Magnusson K, Norén F. Screening of microplastic particles in and down-stream a wastewater treatment plant. Stockholm, Sweden: IVL Swedish Environmental Research Institute; 2014. Report C55.
[32] Talvitie J, Heinonen M, Pääkkönen J-P, et al. Do wastewater treatment plants act as a potential point source of microplastics? Preliminary study in the Coastal Gulf of Finland, Baltic Sea. *Water Sci Technol J Int Assoc Water Pollut Res*. 2015;*72*(9):1495–1504. doi:10.2166/wst.2015.360.
[33] Murphy F, Ewins C, Carbonnier F, Quinn B. Wastewater treatment works (WwTW) as a source of microplastics in the aquatic environment. *Environ Sci Technol*. 2016;*50*(11):5800–5808. doi:10.1021/acs.est.5b05416.
[34] Mintenig SM, Int-Veen I, Löder MGJ, Primpke S, Gerdts G. Identification of microplastic in effluents of waste water treatment plants using focal plane array-based micro-Fourier-transform infrared imaging. *Water Res.* 2017;*108*:365–372. doi:10.1016/j.watres.2016.11.015.
[35] Ziajahromi S, Neale PA, Rintoul L, Leusch FDL. Wastewater treatment plants as a pathway for microplastics: development of a new approach to sample wastewater-based microplastics. *Water Res.* 2017;*112*:93–99. doi:10.1016/j.watres.2017.01.042.
[36] Gies EA, LeNoble JL, Noël M, et al. Retention of microplastics in a major secondary wastewater treatment plant in Vancouver, Canada. *Mar Pollut Bull.* 2018;*133*:553–561. doi:10.1016/j.marpolbul.2018.06.006.
[37] Gündoğdu S, Çevik C, Güzel E, Kilercioğlu S. Microplastics in municipal wastewater treatment plants in Turkey: a comparison of the influent and secondary effluent concentrations. *Environ Monit Assess*. 2018;*190*(11):626. doi:10.1007/s10661-018-7010-y.

[38] Conley K, Clum A, Deepe J, Lane H, Beckingham B. Wastewater treatment plants as a source of microplastics to an urban estuary: removal efficiencies and loading per capita over one year. *Water Res. X*. 2019;*3*:100030. doi:10.1016/j.wroa.2019.100030.
[39] Lv X, Dong Q, Zuo Z, Liu Y, Huang X, Wu W-M. Microplastics in a municipal wastewater treatment plant: fate, dynamic distribution, removal efficiencies, and control strategies. *J Clean Prod*. 2019;*225*:579–586. doi:10.1016/j.jclepro.2019.03.321.
[40] Magni S, Binelli A, Pittura L, et al. The fate of microplastics in an Italian wastewater treatment plant. *Sci. Total Environ*. 2019;*652*:602–610. doi:10.1016/j.scitotenv.2018.10.269.
[41] Leslie HA, Brandsma SH, van Velzen MJM, Vethaak AD. Microplastics en route: field measurements in the Dutch river delta and Amsterdam canals, wastewater treatment plants, North Sea sediments and biota. *Environ Int*. 2017;*101*:133–142. doi:10.1016/j.envint.2017.01.018.
[42] Iyare PU, Ouki SK, Bond T. Microplastics removal in wastewater treatment plants: a critical review. *Environ Sci Water Res Technol*. 2020;*6*(10):2664–2675. doi:10.1039/D0EW00397B.
[43] Dris R, Gasperi J, Rocher V, Saad M, Renault N, Tassin B. Microplastic contamination in an urban area: a case study in Greater Paris. *Environ Chem*. 2015;*12*(5):592–599. doi:10.1071/EN14167.
[44] Lee H, Kim Y. Treatment characteristics of microplastics at biological sewage treatment facilities in Korea. *Mar Pollut Bull*. 2018;*137*:1–8. doi:10.1016/j.marpolbul.2018.09.050.
[45] Simon M, van Alst N, Vollertsen J. Quantification of microplastic mass and removal rates at wastewater treatment plants applying focal plane array (FPA)-based Fourier transform infrared (FT-IR) imaging. *Water Res*. 2018;*142*:1–9. doi:10.1016/j.watres.2018.05.019.
[46] Wisniowska E, Moraczewska-Majkut K, Nocon W. Efficiency of microplastics removal in selected wastewater treatment plants – preliminary studies. *DWT*. 2018;*134*:316–323. doi:10.5004/dwt.2018.23418.
[47] Kleinschmidt JM, Janosik AM. Microplastics in Florida, United States: a case study of quantification and characterization with intertidal snails. *Front Ecol Evol*. 2021;*9*:1–12. doi:10.3389/fevo.2021.645727.
[48] Sobhani Z, Luo Y, Gibson CT, et al. Collecting microplastics in gardens: case study (i) of soil. *Front Environ Sci*. 2021;*9*:1–11. doi:10.3389/fenvs.2021.739775.
[49] Ragusa A, Notarstefano V, Svelato A, et al. Raman microspectroscopy detection and characterisation of microplastics in human breastmilk. *Polymers*. 2022;*14*(13):1–14. doi:10.3390/polym14132700.
[50] Kılıç E, Yücel N, Şahutoğlu SM. Microplastic composition, load and removal efficiency from wastewater treatment plants discharging into Orontes River. *Int J Environ Res*. 2023;*17*(2):25. doi:10.1007/s41742-023-00514-0.
[51] Yang L, Zhang Y, Kang S, Wang Z, Wu C. Microplastics in soil: a review on methods, occurrence, sources, and potential risk. *Sci Total Environ*. 2021;*780*:146546. doi:10.1016/j.scitotenv.2021.146546.
[52] Murray A, Örmeci B. Removal effectiveness of nanoplastics (<400 Nm) with separation processes used for water and wastewater treatment. *Water*. 2020;*12*(3):635. doi:10.3390/w12030635.
[53] Wright SL, Ulke J, Font A, Chan KLA, Kelly FJ. Atmospheric microplastic deposition in an urban environment and an evaluation of transport. *Environ Int*. 2020;*136*:105411. doi:10.1016/j.envint.2019.105411.

[54] Zhang Y, Kang S, Allen S, Allen D, Gao T, Sillanpää M. Atmospheric microplastics: a review on current status and perspectives. *Earth-Science Rev*. 2020;*203*:103118. doi:10.1016/j.earscirev.2020.103118.

[55] Allen S, Allen D, Moss K, Le Roux G, Phoenix VR, Sonke JE. Examination of the ocean as a source for atmospheric microplastics. *PLOS One*. 2020;*15*(5):e0232746. doi:10.1371/journal.pone.0232746.

[56] Botterell ZLR, Beaumont N, Dorrington T, Steinke M, Thompson RC, Lindeque PK. Bioavailability and effects of microplastics on marine zooplankton: a review. *Environ Pollut*. 2019;*245*:98–110. doi:10.1016/j.envpol.2018.10.065.

[57] Lambert S, Scherer C, Wagner M. Ecotoxicity testing of microplastics: considering the heterogeneity of physicochemical properties. *Integr Environ Assess Manag*. 2017;*13*(3):470–475. doi:10.1002/ieam.1901.

[58] Gao H, Yan C, Liu Q, Ding W, Chen B, Li Z. Effects of plastic mulching and plastic residue on agricultural production: a meta-analysis. *Sci Total Environ*. 2019;*651*(Pt 1):484–492. doi:10.1016/j.scitotenv.2018.09.105.

[59] Schwinghammer L, Krause S, Schaum C. Determination of large microplastics: wet-sieving of dewatered digested sludge, co-substrates, and compost. *Water Sci Technol*. 2020;*84*(2):384–392. doi:10.2166/wst.2020.582.

[60] McGivney E, Cederholm L, Barth A, et al. Rapid physicochemical changes in microplastic induced by biofilm formation. *Front Bioeng Biotechnol*. 2020;*8*:1–14. doi:10.3389/fbioe.2020.00205.

[61] Gandara E Silva PP, Nobre CR, Resaffe P, Pereira CDS, Gusmão F. Leachate from microplastics impairs larval development in brown mussels. *Water Res*. 2016;*106*:364–370. doi:10.1016/j.watres.2016.10.016.

[62] O'Connor IA, Golsteijn L, Hendriks AJ. Review of the partitioning of chemicals into different plastics: consequences for the risk assessment of marine plastic debris. *Mar Pollut Bull*. 2016;*113*(1–2):17–24. doi:10.1016/j.marpolbul.2016.07.021.

[63] Hale RC, Seeley ME, La Guardia MJ, Mai L, Zeng EY. A global perspective on microplastics. *JGR Oceans*. 2020;*125*(1):1–40. doi:10.1029/2018JC014719.

[64] Dobslaw D, Woiski C, Kiel M, Kuch B, Breuer J. Plant uptake, translocation and metabolism of PBDEs in plants of food and feed industry: a review. *Rev Environ Sci Biotechnol*. 2021;*20*(1):75–142. doi:10.1007/s11157-020-09557-7.

[65] Engwa GA, Ferdinand PU, Nwalo FN, Unachukwu MN. In: Karcioglu O, Arslan B, eds. *Mechanism and Health Effects of Heavy Metal Toxicity in Humans*. Rijeka: IntechOpen; 2019: 1–13.

[66] Verla AW, Enyoh CE, Verla EN, Nwarnorh KO. Microplastic–toxic chemical interaction: a review study on quantified levels, mechanism and implication. *SN Appl Sci*. 2019;*1*(11):1400. doi:10.1007/s42452-019-1352-0.

[67] Amelia TSM, Khalik WMAWM, Ong MC, Shao YT, Pan H-J, Bhubalan K. Marine microplastics as vectors of major ocean pollutants and its hazards to the marine ecosystem and humans. *Prog Earth Planet Sci*. 2021;*8*(1):12. doi:10.1186/s40645-020-00405-4.

[68] Ma H, Pu S, Liu S, Bai Y, Mandal S, Xing B. Microplastics in aquatic environments: toxicity to trigger ecological consequences. *Environ Pollut*. 2020;*261*:114089. doi:10.1016/j.envpol.2020.114089.

[69] Li J, Zhang K, Zhang H. Adsorption of antibiotics on microplastics. *Environ Pollut*. 2018;*237*:460–467. doi:10.1016/j.envpol.2018.02.050.

[70] Rahman A, Sarkar A, Yadav OP, Achari G, Slobodnik J. Potential human health risks due to environmental exposure to nano- and microplastics and knowledge gaps: a scoping review. *Sci Total Environ*. 2021;*757*:143872. doi:10.1016/j.scitotenv.2020.143872.

[71] Carbery M, O'Connor W, Palanisami T. Trophic transfer of microplastics and mixed contaminants in the marine food web and implications for human health. *Environ Int.* 2018;*115*:400–409. doi:10.1016/j.envint.2018.03.007.
[72] Lehner R, Weder C, Petri-Fink A, Rothen-Rutishauser B. Emergence of nanoplastic in the environment and possible impact on human health. *Environ Sci Technol.* 2019;*53*(4):1748–1765. doi:10.1021/acs.est.8b05512.
[73] Mason SA, Welch VG, Neratko J. Synthetic polymer contamination in bottled water. *Front Chem*. 2018;*6*:1–11. doi:10.3389/fchem.2018.00407.
[74] Ge H, Yan Y, Wu D, Huang Y, Tian F. Potential Role of LINC00996 in Colorectal Cancer: A Study Based on Data Mining and Bioinformatics. *Onco. Targets. Ther*. 2018;*11*:4845–4855. doi:10.2147/OTT.S173225.
[75] Yee MS-L, Hii L-W, Looi CK, et al. Impact of microplastics and nanoplastics on human health. *Nanomater.* 2021;*11*(2):1–22. doi:10.3390/nano11020496.
[76] Stapleton PA. Toxicological considerations of nano-sized plastics. *AIMS Environ Sci*. 2019;*6*(5):367–378. doi:10.3934/environsci.2019.5.367.
[77] Hernandez LM, Yousefi N, Tufenkji N. Are there nanoplastics in your personal care products? *Environ Sci Technol Lett*. 2017;*4*(7):280–285. doi:10.1021/acs.estlett.7b00187.
[78] Dawson A, Huston W, Kawaguchi S, et al. Uptake and depuration kinetics influence microplastic bioaccumulation and toxicity in Antarctic Krill (*Euphausia Superba*). *Environ Sci. Technol.* 2018;*52*(5):3195–3201. doi:10.1021/acs.est.7b05759.
[79] Mitrano DM, Wick P, Nowack B. Placing nanoplastics in the context of global plastic pollution. *Nat Nanotechnol.* 2021;*16*(5):491–500. doi:10.1038/s41565-021-00888-2.
[80] Clark NJ, Khan FR, Mitrano DM, Boyle D, Thompson RC. Demonstrating the translocation of nanoplastics across the fish intestine using palladium-doped polystyrene in a salmon gut-sac. *Environ. Int.* 2022;*159*:106994. doi:10.1016/j.envint.2021.106994.
[81] McIlwraith HK, Kim J, Helm P, Bhavsar SP, Metzger JS, Rochman CM. Evidence of microplastic translocation in wild-caught fish and implications for microplastic accumulation dynamics in food webs. *Environ Sci Technol.* 2021;*55*(18):12372–12382. doi:10.1021/acs.est.1c02922.
[82] So MWK, Vorsatz LD, Cannicci S, Not C. Fate of plastic in the environment: from macro to nano by macrofauna. *Environ Pollut.* 2022;*300*:118920. doi:10.1016/j.envpol.2022.118920.
[83] Fournier SB, D'Errico JN, Adler DS, et al. Nanopolystyrene translocation and fetal deposition after acute lung exposure during late-stage pregnancy. *Part Fibre Toxicol.* 2020;*17*(1):55. doi:10.1186/s12989-020-00385-9.
[84] Hu L, Chernick M, Lewis AM, Ferguson PL, Hinton DE. Chronic microfiber exposure in adult Japanese Medaka (*Oryzias Latipes*). *PLOS One.* 2020;*15*(3):e0229962. doi:10.1371/journal.pone.0229962.
[85] Hu L, Zhou Y, Wang Y, Zhang D, Pan X. Transfer of micro(nano)plastics in animals: a mini-review and future research recommendation. *J Hazard Mater Adv*. 2022;*7*:100101. doi:10.1016/j.hazadv.2022.100101.
[86] Driedger AGJ, Dürr HH, Mitchell K, Van Cappellen P. Plastic debris in the Laurentian Great Lakes: a review. *J Great Lakes Res*. 2015;*41*(1):9–19. doi:10.1016/j.jglr.2014.12.020.
[87] Long M, Moriceau B, Gallinari M, et al. Interactions between microplastics and phytoplankton aggregates : impact on their respective fates. *Mar Chem*. 2015;*175*:39–46. doi:10.1016/j.marchem.2015.04.003.
[88] Musa IO, Auta HS, Ilyasu US, et al. Micro- and nanoplastics in environment: degradation, detection, and ecological impact. *Int J Environ Res*. 2023;*18*(1):1. doi:10.1007/s41742-023-00551-9.

[89] Zantis LJ, Carroll EL, Nelms SE, Bosker T. Marine mammals and microplastics: a systematic review and call for standardisation. *Environ Pollut.* 2021;*269*:116142. doi:10.1016/j.envpol.2020.116142.

[90] Lu L, Luo T, Zhao Y, Cai C, Fu Z, Jin Y. Interaction between microplastics and microorganism as well as gut microbiota: a consideration on environmental animal and human health. *Sci Total Environ.* 2019;*667*:94–100. doi:10.1016/j.scitotenv.2019.02.380.

[91] Cole M, Lindeque P, Fileman E, et al. Microplastic ingestion by zooplankton. *Environ. Sci. Technol.* 2013;*47*(12):6646–6655. doi:10.1021/es400663f.

[92] He M, Yan M, Chen X, et al. Bioavailability and toxicity of microplastics to zooplankton. *Gondwana Res.* 2022;*108*:120–126. doi:10.1016/j.gr.2021.07.021.

[93] EEA. Plastic in textiles : towards a circular economy for synthetic textiles in Europe. https://www.eea.europa.eu/publications/plastic-in-textiles-towards-a. European Environment Agency| Briefing No 25/2020. Published Month June 29, 2021

[94] Romera-Castillo C, Pinto M, Langer TM, Álvarez-Salgado XA, Herndl GJ. Dissolved organic carbon leaching from plastics stimulates microbial activity in the ocean. *Nat Commun.* 2018;*9*(1):1430. doi:10.1038/s41467-018-03798-5.

[95] Miao L, Wang P, Hou J, et al. Distinct community structure and microbial functions of biofilms colonizing microplastics. *Sci Total Environ.* 2019;*650*(Pt 2):2395–2402. doi:10.1016/j.scitotenv.2018.09.378.

[96] McCormick AR, Hoellein TJ, London MG, Hittie J, Scott JW, Kelly JJ. Microplastic in surface waters of urban rivers: concentration, sources, and associated bacterial assemblages. *Ecosphere.* 2016;*7*(11):e01556. doi:10.1002/ecs2.1556.

[97] Gong M, Yang G, Zhuang L, Zeng EY. Microbial biofilm formation and community structure on low-density polyethylene microparticles in lake water microcosms. *Environ Pollut.* 2019;*252*(Pt A):94–102. doi:10.1016/j.envpol.2019.05.090.

[98] Smith M, Love DC, Rochman CM, Neff RA. Microplastics in seafood and the implications for human health. *Curr Environ Heal Reports.* 2018;*5*(3):375–386. doi:10.1007/s40572-018-0206-z.

[99] Hitchcock JN. Microplastics can alter phytoplankton community composition. *Sci Total Environ.* 2022;*819*(January):153074. doi:10.1016/j.scitotenv.2022.153074.

[100] Mao Y, Ai H, Chen Y, et al. Phytoplankton response to polystyrene microplastics: perspective from an entire growth period. *Chemosphere.* 2018;*208*:59–68. doi:10.1016/j.chemosphere.2018.05.170.

[101] Senavirathna MDHJ, Zhaozhi L, Fujino T. Root adsorption of microplastic particles affects the submerged freshwater macrophyte *Egeria Densa*. *Water Air Soil Pollut.* 2022;*233*(3):80. doi:10.1007/s11270-022-05556-2.

[102] Zhang C, Jian M-F, Chen Y-M[, et al. Effects of polystyrene microplastics (PS-MPs) on the growth, physiology, and biochemical characteristics of *Hydrilla verticillata*. *Ying Yong Sheng Tai Xue Bao* [J Appl Ecol]. 2021;*32*(1):317–325. doi:10.13287/j.1001-9332.202101.037.

[103] Klein M, Fischer EK. Microplastic abundance in atmospheric deposition within the metropolitan area of Hamburg, Germany. *Sci Total Environ.* 2019;*685*:96–103. doi:10.1016/j.scitotenv.2019.05.405.

[104] Baensch-Baltruschat B, Kocher B, Stock F, Reifferscheid G. Tyre and road wear particles (TRWP) – a review of generation, properties, emissions, human health risk, ecotoxicity, and fate in the environment. *Sci Total Environ.* 2020;*733*:137823. doi:10.1016/j.scitotenv.2020.137823.

[105]SAPEA, Science Advice for Policy By European Academies. *A Scientific Perspective on Microplastics in Nature and Society*. Berlin: SAPEA; 2019; Evidence Review Report No. 4: 1-173. doi: 10.26356/microplastics .

[106]Henry B, Laitala K, Klepp IG. Microfibres from apparel and home textiles: prospects for including microplastics in environmental sustainability assessment. *Sci Total Environ.* 2019;*652*:483–494. doi:10.1016/j.scitotenv.2018.10.166.

[107]Gong J, Xie P. Research progress in sources, analytical methods, eco-environmental effects, and control measures of microplastics. *Chemosphere.* 2020;*254*:126790. doi:10.1016/j.chemosphere.2020.126790.

[108]de Souza Machado AA, Lau CW, Till J, et al. Impacts of microplastics on the soil biophysical environment. *Environ Sci Technol.* 2018;*52*(17):9656–9665. doi:10.1021/acs.est.8b02212.

[109]de Souza Machado AA, Lau CW, Kloas W, et al. Microplastics can change soil properties and affect plant performance. *Environ Sci Technol.* 2019;*53*(10):6044–6052. doi:10.1021/acs.est.9b01339.

[110]Wang J, Huang M, Wang Q, Sun Y, Zhao Y, Huang Y. LDPE microplastics significantly alter the temporal turnover of soil microbial communities. *Sci Total Environ.* 2020;*726*:138682. doi:10.1016/j.scitotenv.2020.138682.

[111]Rillig MC, de Souza Machado AA, Lehmann A, Klümper U. Evolutionary implications of microplastics for soil biota. *Environ Chem.* 2019;*16*(1):3–7. doi:10.1071/EN18118.

[112]Zhang Y, Zhang X, Li X, He D. Interaction of microplastics and soil animals in agricultural ecosystems. *Curr Opin Environ Sci Heal.* 2022;*26*:100327. doi:10.1016/j.coesh.2022.100327.

[113]Ding J, Liu C, Chen Q, et al. Extractable additives in microplastics: a hidden threat to soil fauna. *Environ Pollut.* 2022;*294*:118647. doi:10.1016/j.envpol.2021.118647.

[114]Liu K, Wang X, Song Z, Wei N, Li D. Terrestrial plants as a potential temporary sink of atmospheric microplastics during transport. *Sci Total Environ.* 2020;*742*:140523. doi:10.1016/j.scitotenv.2020.140523.

[115]Luo X, Bing H, Luo Z, Wang Y, Jin L. Impacts of atmospheric particulate matter pollution on environmental biogeochemistry of trace metals in soil-plant system: a review. *Environ Pollut.* 2019;*255*:113138. doi:10.1016/j.envpol.2019.113138.

[116]Lian J, Liu W, Meng L, et al. Foliar-applied polystyrene nanoplastics (PSNPs) reduce the growth and nutritional quality of lettuce (*Lactuca Sativa* L.). *Environ Pollut.* 2021;*280*:116978. doi:10.1016/j.envpol.2021.116978.

[117]Meng F, Yang X, Riksen M, Geissen V. Effect of different polymers of microplastics on soil organic carbon and nitrogen – a mesocosm experiment. *Environ Res.* 2022;*204*:111938. doi:10.1016/j.envres.2021.111938.

[118]Huang D, Wang X, Yin L, et al. Research progress of microplastics in soil-plant system: ecological effects and potential risks. *Sci Total Environ.* 2022;*812*:151487. doi:10.1016/j.scitotenv.2021.151487.

[119]Wang Y, Huang J, Zhu F, Zhou S. Airborne microplastics: a review on the occurrence, migration and risks to humans. *Bull Environ Contam Toxicol.* 2021;*107*(4):657–664. doi:10.1007/s00128-021-03180-0.

[120]Vianello A, Jensen RL, Liu L, Vollertsen J. Simulating human exposure to indoor airborne microplastics using a breathing thermal manikin. *Sci Rep*. 2019;*9*(1):8670. doi:10.1038/s41598-019-45054-w.

[121]Alderete TL, Jones RB, Chen Z, et al. Exposure to traffic-related air pollution and the composition of the gut microbiota in overweight and obese adolescents. *Environ Res.* 2018;*161*:472–478. doi:10.1016/j.envres.2017.11.046.

[122]Soltani NS, Taylor MP, Wilson SP. Quantification and exposure assessment of microplastics in Australian indoor house dust. *Environ Pollut.* 2021;*283*:117064. doi:10.1016/j.envpol.2021.117064.

[123]Shi W, Cao Y, Chai X, et al. Potential health risks of the interaction of microplastics and lung surfactant. *J Hazard Mater.* 2022;*429*:128109. doi:10.1016/j.jhazmat.2021.128109.

[124]Goodman KE, Hare JT, Khamis ZI, Hua T, Sang Q-XA. Exposure of human lung cells to polystyrene microplastics significantly retards cell proliferation and triggers morphological changes. *Chem Res Toxicol.* 2021;*34*(4):1069–1081. doi:10.1021/acs.chemrestox.0c00486.

[125]Xu M, Halimu G, Zhang Q, et al. Internalization and toxicity: a preliminary study of effects of nanoplastic particles on human lung epithelial cell. *Sci Total Environ.* 2019;*694*:133794. doi:10.1016/j.scitotenv.2019.133794.

[126]Dong C-D, Chen C-W, Chen Y-C, Chen H-H, Lee J-S, Lin C-H. Polystyrene microplastic particles: *in vitro* pulmonary toxicity assessment. *J Hazard Mater.* 2020;*385*:121575. doi:10.1016/j.jhazmat.2019.121575.

[127]Zhang J, Wang L, Trasande L, Kannan K. Occurrence of polyethylene terephthalate and polycarbonate microplastics in infant and adult feces. *Environ Sci Technol Lett.* 2021;*8*(11):989–994. doi:10.1021/acs.estlett.1c00559.

[128]Fackelmann G, Sommer S. Microplastics and the gut microbiome: how chronically exposed species may suffer from gut dysbiosis. *Mar Pollut Bull.* 2019;*143*:193–203. doi:10.1016/j.marpolbul.2019.04.030.

[129]Stock V, Fahrenson C, Thuenemann A, et al. Impact of artificial digestion on the sizes and shapes of microplastic particles. *Food Chem Toxicol.* 2020;*135*:111010. doi:10.1016/j.fct.2019.111010.

[130]Huang W, Yin H, Yang Y, Jin L, Lu G, Dang Z. Influence of the co-exposure of microplastics and tetrabromobisphenol A on human gut: simulation *in vitro* with human cell Caco-2 and gut microbiota. *Sci Total Environ.* 2021;*778*:146264. doi:10.1016/j.scitotenv.2021.146264.

[131]Stock V, Böhmert L, Lisicki E, et al. Uptake and effects of orally ingested polystyrene microplastic particles *in vitro* and *in vivo*. *Arch Toxicol.* 2019;*93*(7):1817–1833. doi:10.1007/s00204-019-02478-7.

[132]Deng Y, Zhang Y, Lemos B, Ren H. Tissue accumulation of microplastics in mice and biomarker responses suggest widespread health risks of exposure. *Sci Rep.* 2017;*7*(1):46687. doi:10.1038/srep46687.

[133]Ibrahim YS, Tuan Anuar S, Azmi AA, et al. Detection of microplastics in human colectomy specimens. *JGH Open.* 2021;*5*(1):116–121. doi:10.1002/jgh3.12457.

[134]Tan M, Sun Y, Gui J, Wang J, Chen X, Wu D. The efficiency of different digestion and separation methods for extracting microplastics in typical organic solid waste. *Int J Environ Res.* 2022;*16*(2):16. doi:10.1007/s41742-022-00394-w.

[135]Heo NW, Hong SH, Han GM, et al. Distribution of small plastic debris in cross-section and high strandline on Heungnam Beach, South Korea. *Ocean Sci J.* 2013;*48*(2):225–233. doi:10.1007/s12601-013-0019-9.

[136]Hidalgo-Ruz V, Thiel M. Distribution and abundance of small plastic debris on beaches in the SE Pacific (Chile): a study supported by a citizen science project. *Mar Environ Res.* 2013;*87–88*:12–18. doi:10.1016/j.marenvres.2013.02.015.

[137]Desforges J-PW, Galbraith M, Dangerfield N, Ross PS. Widespread distribution of microplastics in subsurface seawater in the NE Pacific Ocean. *Mar. Pollut. Bull.* 2014;*79*(1–2):94–99. doi:10.1016/j.marpolbul.2013.12.035.

[138]Song YK, Hong SH, Jang M, et al. A comparison of microscopic and spectroscopic identification methods for analysis of microplastics in environmental samples. *Mar Pollut Bull.* 2015;*93*(1–2):202–209. doi:10.1016/j.marpolbul.2015.01.015.
[139]Shim WJ, Hong SH, Eo SE. Identification methods in microplastic analysis: a review. *Anal Methods*. 2017;*9*(9):1384–1391. doi:10.1039/C6AY02558G.
[140]Song YK, Hong SH, Jang M, et al. Large accumulation of micro-sized synthetic polymer particles in the sea surface microlayer. *Environ. Sci. Technol.* 2014;*48*(16):9014–9021. doi:10.1021/es501757s.
[141]Han XX, Rodriguez RS, Haynes CL, Ozaki Y, Zhao B. Surface-enhanced raman spectroscopy. *Nat Rev Methods Prim*. 2022;*1*(1):87. doi:10.1038/s43586-021-00083-6.
[142]Mogha NK, Shin D. Nanoplastic detection with surface enhanced Raman spectroscopy: present and future. *TrAC Trends Anal Chem*. 2023;*158*:116885. doi:10.1016/j.trac.2022.116885.
[143]Li P, Long F, Chen W, Chen J, Chu PK, Wang H. Fundamentals and applications of surface-enhanced raman spectroscopy–based biosensors. *Curr Opin Biomed Eng*. 2020;*13*:51–59. doi:10.1016/j.cobme.2019.08.008.
[144]Mandal P, Tewari BS. Progress in surface enhanced Raman scattering molecular sensing: a review. *Surf Interf.* 2022;*28*:101655. doi:10.1016/j.surfin.2021.101655.
[145]Yılmaz D, Günaydın BN, Yüce M. Nanotechnology in food and water security: on-site detection of agricultural pollutants through surface-enhanced Raman spectroscopy. *Emergent Mater*. 2022;*5*(1):105–132. doi:10.1007/s42247-022-00376-w.
[146]Jiménez-Lamana J, Marigliano L, Allouche J, Grassl B, Szpunar J, Reynaud S. A novel strategy for the detection and quantification of nanoplastics by single particle inductively coupled plasma mass spectrometry (ICP-MS). *Anal Chem*. 2020;*92*(17):11664–11672. doi:10.1021/acs.analchem.0c01536.
[147]Lai Y, Dong L, Li Q, et al. Counting nanoplastics in environmental waters by single particle inductively coupled plasma mass spectroscopy after cloud-point extraction and *in situ* labeling of gold nanoparticles. *Environ Sci Technol.* 2021;*55*(8):4783–4791. doi:10.1021/acs.est.0c06839.
[148]Laborda F, Trujillo C, Lobinski R. Analysis of microplastics in consumer products by single particle-inductively coupled plasma mass spectrometry using the carbon-13 isotope. *Talanta.* 2021;*221*:121486. doi:10.1016/j.talanta.2020.121486.
[149]Randhawa JS. Advanced analytical techniques for microplastics in the environment: a review. *Bull Natl Res Cent*. 2023;*47*(1):174. doi:10.1186/s42269-023-01148-0.
[150]Okoffo ED, Ribeiro F, O'Brien JW, et al. Identification and quantification of selected plastics in biosolids by pressurized liquid extraction combined with double-shot pyrolysis gas chromatography–mass spectrometry. *Sci Total Environ.* 2020;*715*:136924. doi:10.1016/j.scitotenv.2020.136924.
[151]Blomberg J. The Mechanism Behind Internal Injector Deposits [Thesis] Stockholm, Sweden: KTH Royal Institute of Technology; 2023
[152]Peng C, Tang X, Gong X, Dai Y, Sun H, Wang L. Development and application of a mass spectrometry method for quantifying nylon microplastics in environment. *Anal. Chem.* 2020;*92*(20):13930–13935. doi:10.1021/acs.analchem.0c02801.
[153]Zhang J, Wang L, Kannan K. Quantitative analysis of polyethylene terephthalate and polycarbonate microplastics in sediment collected from South Korea, Japan and the USA. *Chemosphere.* 2021;*279*:130551. doi:10.1016/j.chemosphere.2021.130551.
[154]Ng KT, Rapp-Wright H, Egli M, et al. High-throughput multi-residue quantification of contaminants of emerging concern in wastewaters enabled using direct injection liquid chromatography-tandem mass spectrometry. *J Hazard Mater.* 2020;*398*:122933. doi:10.1016/j.jhazmat.2020.122933.

[155] Lin Y, Huang X, Liu Q, Lin Z, Jiang G. Thermal fragmentation enhanced identification and quantification of polystyrene micro/nanoplastics in complex media. *Talanta*. 2020;*208*:120478. doi:10.1016/j.talanta.2019.120478.

[156] Adhikari S, Kelkar V, Kumar R, Halden RU. Methods and challenges in the detection of microplastics and nanoplastics: a mini-review. *Polym Int*. 2022;*71*(5):543–551. doi:10.1002/pi.6348.

[157] Moon S, Martin LMA, Kim S, et al. Direct observation and identification of nanoplastics in ocean water. *Sci Adv*. 2024;*10*(4):eadh1675. doi:10.1126/sciadv.adh1675.

[158] Tse Y-T, Lo H-S, Chan SM, Sze ET. Flow cytometry as a rapid alternative to quantify small microplastics in environmental water samples. *Water*. 2022;*14*(9):1436. doi:10.3390/w14091436.

[159] Kaile N, Lindivat M, Elio J, Thuestad G, Crowley QG, Hoell IA. Preliminary results from detection of microplastics in liquid samples using flow cytometry. *Front Mar Sci*. 2020;*7*: 1–12. doi:10.3389/fmars.2020.552688.

[160] Pérez-Guevara F, Roy PD, Kutralam-Muniasamy G, Shruti VC. A central role for fecal matter in the transport of microplastics: an updated analysis of new findings and persisting questions. *J Hazard Mater Adv*. 2021;*4*:100021. doi:10.1016/j.hazadv.2021.100021.

[161] Mansa R, Zou S. Thermogravimetric analysis of microplastics: a mini review. *Environ Adv*. 2021;*5*:100117. doi:10.1016/j.envadv.2021.100117.

[162] Rai PK, Kumar V, Sonne C, Lee SS, Brown RJC, Kim K-H. Progress, prospects, and challenges in standardization of sampling and analysis of micro- and nano-plastics in the environment. *J Clean Prod*. 2021;*325*:129321. doi:10.1016/j.jclepro.2021.129321.

[163] Ho VTTX, Park H, An S, et al. Coumarin–lipoic acid conjugates on silver nanoparticle-supported nanopipettes for *in situ* dual-mode monitoring of intracellular Cu(II) and potential chemodynamic therapy applications. *Sensors Actuators B Chem*. 2021;*344*:130271. doi:10.1016/j.snb.2021.130271.

[164] Nie X-L, Liu H-L, Pan Z-Q, et al. Recognition of plastic nanoparticles using a single gold nanopore fabricated at the tip of a glass nanopipette. *Chem Commun*. 2019;*55*(45):6397–6400. doi:10.1039/C9CC01358J.

[165] Boni W, Arbuckle-Keil G, Fahrenfeld NL. Inter-storm variation in microplastic concentration and polymer type at stormwater outfalls and a bioretention basin. *Sci Total Environ*. 2022;*809*:151104. doi:10.1016/j.scitotenv.2021.151104.

[166] Balestra V, Bellopede R. Microplastic pollution in show cave sediments: first evidence and detection technique. *Environ Pollut*. 2022;*292*:118261. doi:10.1016/j.envpol.2021.118261.

[167] Rist S, Hartmann NB, Welden NAC. How fast, how far: diversification and adoption of novel methods in aquatic microplastic monitoring. *Environ Pollut*. 2021;*291*:118174. doi:10.1016/j.envpol.2021.118174.

[168] Elsayed AA, Erfan M, Sabry YM, et al. A microfluidic chip enables fast analysis of water microplastics by optical spectroscopy. *Sci Rep*. 2021;*11*(1):10533. doi:10.1038/s41598-021-89960-4.

[169] Matthews C, Moran F, Jaiswal AK. A review on European Union's strategy for plastics in a circular economy and its impact on food safety. *J Clean Prod*. 2021;*283*:125263. doi:10.1016/j.jclepro.2020.125263.

[170] Mitrano DM, Wohlleben W. Microplastic regulation should be more precise to incentivize both innovation and environmental safety. *Nat Commun*. 2020;*11*(1):5324. doi:10.1038/s41467-020-19069-1.

[171] Puharinen S-T. Achieving good marine environmental status in the EU – implications of the marine strategy framework directive for member states and blue economic activities. *Mar Policy*. 2023;*155*:105712. doi:10.1016/j.marpol.2023.105712.
[172] Anagnosti L, Varvaresou A, Pavlou P, Protopapa E, Carayanni V. Worldwide actions against plastic pollution from microbeads and microplastics in cosmetics focusing on European policies. Has the issue been handled effectively? *Mar Pollut Bull*. 2021;*162*:111883. doi:10.1016/j.marpolbul.2020.111883.
[173] Alava JJ, Lukyanova ON, Ross PS, Shim WJ.. Report of Working Group 31 on Emerging Topics in Marine Pollution. Canada: North Pacific Marine Science Organization (PICES); 2020. PICES Scientific Report No. 56
[174] Akhbarizadeh R, Yu JT, Ead L, et al. Reductions of plastic microbeads from personal care products in wastewater effluents and lake waters following regulatory actions. *ACS ES&T Water*. 2024;*4*(2):492–499. doi:10.1021/acsestwater.3c00526.
[175] Wuennenberg L, Tan CM. Plastic waste in Canada. Canada: International Institute for Sustainable Development; 2019
[176] Hu J, Miao L, Han J, Zhou W, Qian X. Waste separation behavior with a new plastic category for the plastic resource circulation: survey in Kansai, Japan. *J Environ Manage*. 2024;*349*:119370. doi:10.1016/j.jenvman.2023.119370.
[177] Agamuthu P, Mehran SB, Norkhairah A, Norkhairiyah A. Marine debris: a review of impacts and global initiatives. *Waste Manag. Res.* 2019;*37*(10):987–1002. doi:10.1177/0734242X19845041.
[178] Shooshtarian S, Maqsood T, Wong PSP, Yang RJ, Khalfan M. Review of waste strategy documents in Australia: analysis of strategies for construction and demolition waste. *IJETM*. 2020;*23*(1):1–21. doi:10.1504/IJETM.2020.110147.
[179] Sun Y, Liu S, Wang P, Jian X, Liao X, Chen W-Q. China's roadmap to plastic waste management and associated economic costs. *J Environ Manage*. 2022;*309*:114686. doi:10.1016/j.jenvman.2022.114686.
[180] Meng M, Wen Z, Luo W, Wang S. Approaches and policies to promote zero-waste city construction: China's practices and lessons. *Sustainability*. 2021;*13*(24):13537. doi:10.3390/su132413537.
[181] Venus O. National marine microbead policy in developed nations: how microbead bans have influenced microplastic pollution in waterways and begun the trend towards international collaboration [Thesis]. Ann Arbor, US: University of Michigan; 2020.
[182] Lopez EMB. Marine debris removal : a case study on South Korean marine litter management [Thesis]. Kingston, US: University of Rhode Island; 2021.
[183] Carlini G, Kleine K. Advancing the international regulation of plastic pollution beyond the United Nations environment assembly resolution on marine litter and microplastics. *Rev Euro Comp Intl Enviro*. 2018;*27*(3):234–244. doi:10.1111/reel.12258.
[184] Borg S, Attard FG, Vella de Fremeaux PM. *Research Handbook on Ocean Governance Law*. Cheltenham, UK: Edward Elgar Publishing; 2023.
[185] Synn K, Thesis JM. Strengthening global governance on marine plastic litter. 2020.

Journal of Applied Physics

RESEARCH ARTICLE | JULY 10 2024

Quadrupole mode plasmon resonance enabled dual-band metamaterial for refractive index sensing application

Ankit ; Monu Nath Baitha ; Kamal Kishor ; Ravindra Kumar Sinha

Check for updates

J. Appl. Phys. 136, 023104 (2024)
https://doi.org/10.1063/5.0201422

View Online
Export Citation

18 July 2024 04:56:20

AIP Publishing

Journal of Applied Physics
Special Topic:
Phonon-Magnon Interactions:
From Fundamentals to Device Physics
Guest Editors: Vasily V. Temnov, Alexey V. Scherbakov, Yoshi Chika Otani, Paolo Vavassori
Submit Today!
AIP Publishing

Journal of Applied Physics ARTICLE pubs.aip.org/aip/jap

Quadrupole mode plasmon resonance enabled dual-band metamaterial for refractive index sensing application

Cite as: J. Appl. Phys. **136**, 023104 (2024); doi: 10.1063/5.0201422
Submitted: 30 January 2024 · Accepted: 22 June 2024 ·
Published Online: 10 July 2024

View Online Export Citation CrossMark

Ankit,[1] Monu Nath Baitha,[2] Kamal Kishor,[1] and Ravindra Kumar Sinha[1,3,a)]

AFFILIATIONS

[1]TIFAC—Center of Relevance and Excellence in Fiber Optics and Optical Communication, Department of Applied Physics, Delhi Technological University, Delhi 110042, India
[2]School of Mechanical Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[3]Gautam Buddha University, Gautam Budh Nagar, Greater Noida, UP 201312, India

[a)]**Author to whom correspondence should be addressed:** dr_rk_sinha@yahoo.com

ABSTRACT

In this paper, design and fabrication of a dual-band near-zero index metamaterial (MTM) structure using copper on an epoxy resin fiber (FR-4) dielectric substrate is reported for refractive index sensing applications. The primary objective is to achieve dual-band operation spanning a 1–15 GHz frequency range, with a specific focus on achieving a broad bandwidth in the C-band. The resonance of the MTM structure was ascribed to the coupling of plane electromagnetic waves with surface plasmon polaritons on the structure, resulting in a quadrupole plasmon resonance mode. Furthermore, transmission characteristics of the fabricated MTM structure were experimentally measured and found to align closely with the simulated results obtained through the finite element method in COMSOL Multiphysics. The designed MTM structure demonstrates negative and near-zero permittivity at resonance frequencies, enabling left-handed and near-zero index behavior in dual microwave frequency bands. Under room temperature conditions, the MTM sensor exhibited sensitivities of 1 GHz/RIU and 3 GHz/RIU at resonance frequencies of 2.7 and 7.3 GHz, respectively. Consequently, the MTM structure exhibits significant potential for diverse applications, serving as a valuable component in sensors, detectors, and optoelectronic devices operating in the GHz region.

18 July 2024 04:56:20

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0201422

I. INTRODUCTION

The S (2–4 GHz) band and the C (4–8 GHz) band hold significant importance across a spectrum of applications due to the unique electromagnetic characteristics of these frequency ranges. Modern optoelectronic devices developed for these bands are essential for various purposes, including wireless communications,[1] radar navigation,[2] remote sensing,[3] etc. However, effective utilization of the S and C bands often encounters challenges, primarily associated with the limitations of conventional materials and components in achieving the desired performance levels. These limitations can impede progress in the above-mentioned application and beyond. Metamaterials (MTMs) represent a groundbreaking solution that has revolutionized the world of electromagnetic engineering. These artificially engineered materials are engineered to manipulate electromagnetic waves in unconventional ways that were previously thought impossible.[4–6] One of the most remarkable feats achieved by metamaterials is the creation of a negative refractive index, allowing waves to bend in unconventional directions.[7] However, the evolution of metamaterials extends beyond this milestone, delving into the realm of zero-index materials—a cutting-edge development within the field. The zero-index metamaterials (ZIMs), a subset of metamaterials, present unique advantages that render them highly valuable for applications such as electromagnetic cloaking, tunneling, and beam splitting.[8] With a refractive index of zero, these materials exhibit remarkable properties that can significantly enhance the performance and efficiency of devices across various frequency ranges. These capabilities open new avenues for manipulating and controlling electromagnetic waves, promising a new range of advantages and applications that go beyond the capabilities of conventional materials.

© Author(s) 2024

MTMs can be classified according to their effective permittivity ($\varepsilon_{\text{eff.}}$) and effective permeability ($\mu_{\text{eff.}}$) parameters, resulting in categories such as single negative (SNG), double negative, and double positive metamaterials. Within the SNG metamaterials, there are two types known as ϵ-negative (ENG) and μ-negative (MNG) materials, depending on whether $\varepsilon_{\text{eff.}}$ or $\mu_{\text{eff.}}$ is negative.[9] Exploration of metamaterials in the GHz and THz frequency ranges has revealed their distinctive properties such as negative permeability (μ), negative permittivity (ε), and negative refractive index (η). As a result, these MTMs have found diverse applications in microwave and optical domains,[10] ranging from reconfigurable antennas,[11] satellites,[12] and dual-band sensors[13] to applications like invisibility cloaking,[14] tunable perfect reflectors,[15] filters,[16] self-adjustable smart metamaterial waveguides,[17] and super-lenses.[18,19] In recent investigations, there has been extensive exploration of sensors operating in several gigahertz ranges, leveraging MTMs. It has been shown that a relatively straightforward geometric structural design can yield impressive sensing capabilities across various applications. Enhanced sensitivity is crucial in sensing applications as it enables the detection of subtle changes in the surrounding environment. The MTM design has shown its great potential in sensing applications. The specific metamaterial belongs to the targeted sensing application as shown by Banerjee *et al.* SRR type of metasurface showed a sensitivity of 0.742 THz/RIU for ultrathin subwavelength films sensing through excitation of dark modes in the THz region.[20] Navaratna *et al.* showed all silicon topological waveguide cavity chips with a honeycomb lattice design operating in the THz range for biomolecule sensing applications.[21] Gupta and Singh[22] showed that a large Q/V meta-sensor emphasizing a high quality factor and a small mode volume with an I-shaped metasurface which resonates at 0.6 THz has a sensitivity of 18 GHz/RIU. Hyperbolic meta-sensors are known for their high-resolution sensing capabilities. Sreekanth *et al.* showed phase change materials based on low-loss visible frequency hyperbolic metamaterials for ultrasensitive label free biosensing applications.[23] Absorptive liquid meta-sensors are optimized for liquid environments of water methanol mixtures. Chen *et al.* showed high sensitivity due to multiple resonances, i.e., Fano resonance and dipole resonance, in the THz region metamaterial of highly absorptive water–methanol mixtures.[24] Microfluidic meta-sensors integrate fluidic channels for enhanced analyte interaction with dual torus (doughnut-shaped object) toroidal metasurfaces.[25] This designed structure is fabricated using single step lithography and shows a high sensitivity of 0.129 THz/RIU for 28 μm thick microfluidic layers of different mixed ethanol–water solutions. Lim *et al.* reviewed the high-Q planar Fano resonant metamaterial sensors that leverage interference effects to achieve high sensitivity toward futuristic technologies.[26] In this continuous development of metamaterial advancement in sensor applications, our study has shown that the MTM sensor utilizes quadrupole plasmon resonance to achieve high sensitivity in dual frequency bands in the GHz range of the S and C bands of microwave regions. This distinct mechanism not only provides high sensitivity but also ensures broad bandwidth coverage in the GHz range, which is a unique feature in the advancement of refractive index sensing. In 2017, Tian *et al.* developed an SRR configuration to create a refractive index sensor achieving a sensitivity of up to 1.69 GHz/RIU in the C-band.[54] Meanwhile, Hu and Liu[29] devised a microwave resonator measuring 18×18 mm^2 but failed to achieve significant sensitivity in the C-band. Additionally, Dadouche *et al.* obtained a corona-shaped metamaterial resonator for the S-band with low sensitivity in 2023.[55] Nevertheless, existing literature has identified constraints in these described MTM structures, underscoring the need for an enhanced dual-band MTM tunable sensor structure capable of achieving high sensitivities in both S- and C-bands.

Over the past several years, numerous MTM structures including V-shaped, modified H-shaped, oval-shaped, I-shaped, and Z-shaped designs have been developed to cater the increasing demands in antenna design for sensor, satellite, and radar communication applications.[30–34] The main goal has been aimed to improve the antenna's efficiency. Achieving compactness and miniaturization of MTMs relies on enhancing the bandwidth, which is a crucial factor in MTM research. In 2020, Rashedul Islam *et al.*[35] introduced an 8×8 mm^2 square enclosed circle split ring resonator (SRR) designed for satellite communications in the S, C, and X bands. This SRR had bandwidths of 0.11, 0.79, and 0.15 GHz in these respective bands. Similarly, Islam *et al.*[36] introduced an MTM structure designed to operate in the S, C, X, and Ku bands, exhibiting bandwidths of 0.19, 0.14, 1.3, and 0.6 GHz, respectively. Moniruzzaman *et al.*[37] suggested a symmetric resonator for the C, X, and Ku bands with bandwidths of 0.72, 1.55, and 0.24 GHz, respectively. Bellal Hossain *et al.*[38] presented a symmetric hexagonal split ring resonator MTM with dual-band operation and a wide bandwidth of 1.67 in the C-band for Wi-Fi applications. More recently, Hossain *et al.*[39] proposed an MTM structure with a bandwidth of 0.94 GHz in the C-band. However, the literature has shown limitations in these reported MTM structures, highlighting the necessity for an improved dual-band MTM structure with a wide bandwidth in the C-band.

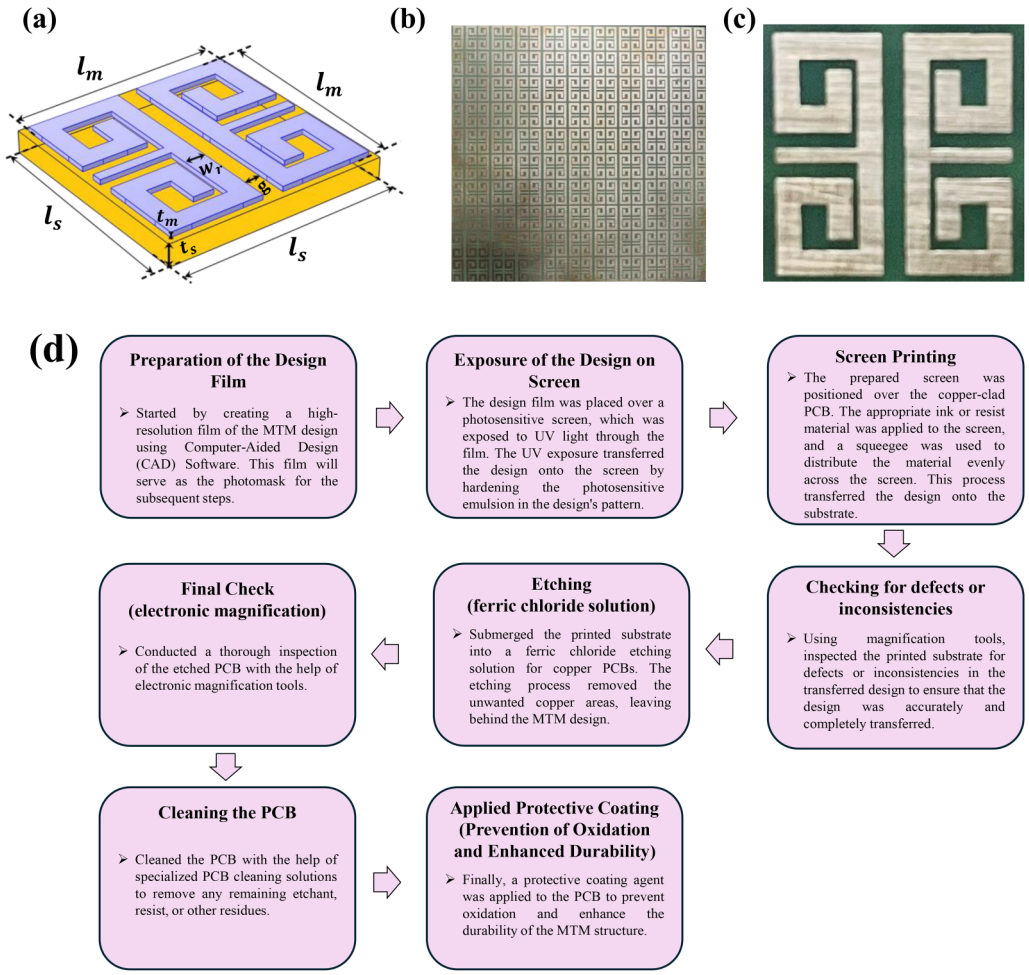
In this current research, we have fabricated and examined a novel metamaterial (MTM) structure designed for dual-band operation, showcasing a broad bandwidth in the C-band for refractive index sensing applications. The MTM unit cell, also called meta-atom's scattering parameter, was studied using FEM based simulation, and the results were validated using experimentally measured results. Analysis of the proposed MTM structure reveals the presence of negative and near-zero effective refractive index (NZI) and epsilon near zero (ENZ) in the dual-band operation range, covering the S and C bands within the microwave region. The designed MTM structure demonstrates a high EMR of up to 6.93, which confirms the compactness of the proposed design. This study suggests that the MTM structure has the potential to significantly enhance the efficiency and compactness of devices and components used in wireless communication systems with its compact size, dual-band resonance, and wide bandwidth in the C band. Utilizing the proposed MTM as a refractive index sensor results in high sensitivity, specifically 1 and 3 GHz/RIU. To create high-performance multifunction sensors with exceptional sensing capabilities, this study presents a viable alternative technique. These sensors would be ideal for use in biochemical sensing, medical detection, and diagnostics.

II. FABRICATION AND MODELING OF THE DESIGNED MTM STRUCTURE

The schematic representation of the meta-atom of the designed MTM structure is demonstrated in Fig. 1(a), which is similarly used for simulation purposes. Figure 1(b) displays the fabricated prototype array of the same, periodically extended on all sides with a periodicity of 16×16 mm^2. The top view of the fabricated meta-atom (unit cell) structure is displayed in Fig. 1(c). The

18 July 2024 04:56:20

© Author(s) 2024



18 July 2024 04:56:20

FIG. 1. (a) The meta-atom or unit cell of the simulated MTM design from an orthogonal view. (b) The physical prototype array of the fabricated MTM design, (c) the top view of the meta-atom of the fabricated sample, and (d) flow chart representing the complete process of fabrication of the simulated MTM design.

designed MTM is composed of a copper (Cu) material resonating patch impinges on a 1.6 mm thick (t_s) epoxy resin fiber substrate. Fabrication of the unit cell of dimension $16 \times 16\ mm^2$ and a prototype array of 10×10 is achieved using a computer numerical control (CNC) machine, shown in Fig. 1(b). The fabrication process begins with creating a design layout using computer-aided design (CAD) or similar software, serving as the blueprint for the MTM structure. Next, the design is transferred onto a screener mask for subsequent printing. Material deposition onto the substrate follows using screen printing, a technique for depositing materials onto substrates. This step involves applying the desired material onto the substrate according to the design pattern. Etching is then performed to refine the geometry of the MTM structure by removing excess materials from the substrate. The final stages entail cleaning the fabricated MTM to eliminate any residues or contaminants remaining from the fabrication process. All the above fabrication processes are presented using a flow chart as shown in Fig. 1(d). All necessary physical parameters are listed in Table I.

TABLE I. The parameters of the fabricated and simulated designed unit cell MTM structure.

Parameter	Value (mm)	Description
l_m	15	Length of the material
G	1.5	Middle gap
l_s	16	Length of the substrate
w_r	1.5	Width of the resonator
t_s	1.6	Thickness of the substrate
t_m	0.35	Thickness of the resonator

The proposed structure is designed and simulated using the FEM based COMSOL Multiphysics simulation software, as

© Author(s) 2024

depicted in Fig. 1(a). The copper material patch in the design has a thickness (t_m) of 0.35 mm with an electrical conductivity (σ) of 5.99×10^7 S/m.[38] The dielectric substrate used in this study has a loss tangent (δ) of 0.025 radians and permittivity (ε) of 4.3.[36,38]

III. RESULTS AND DISCUSSION

A. Scattering parameter of the optimized metamaterial unit cell

Scattering parameters of the simulated design structure were computed using the Drude–Lorentz variables. A configuration with two waveguide ports was used, where port 1 acted as the signal transmitter, and port 2 served as the receiver [Fig. 2(a)]. The x axis was implemented with a perfect magnetic conductor (PMC), while the y axis had a perfect electric conductor (PEC) with periodic boundary conditions in x and y axes. The electromagnetic wave propagated vertically along the z axis. This setup allowed for precise simulations and analysis of the electromagnetic behavior of the proposed design.

The experimental configuration used for S_{21} measurement involves an antenna test setup illustrated in Fig. 2(b). This setup comprises connectors, RF cables, and calibration standards necessary for accurate measurements. A vector network analyzer (VNA-MS2028C) is employed for this purpose. The device under

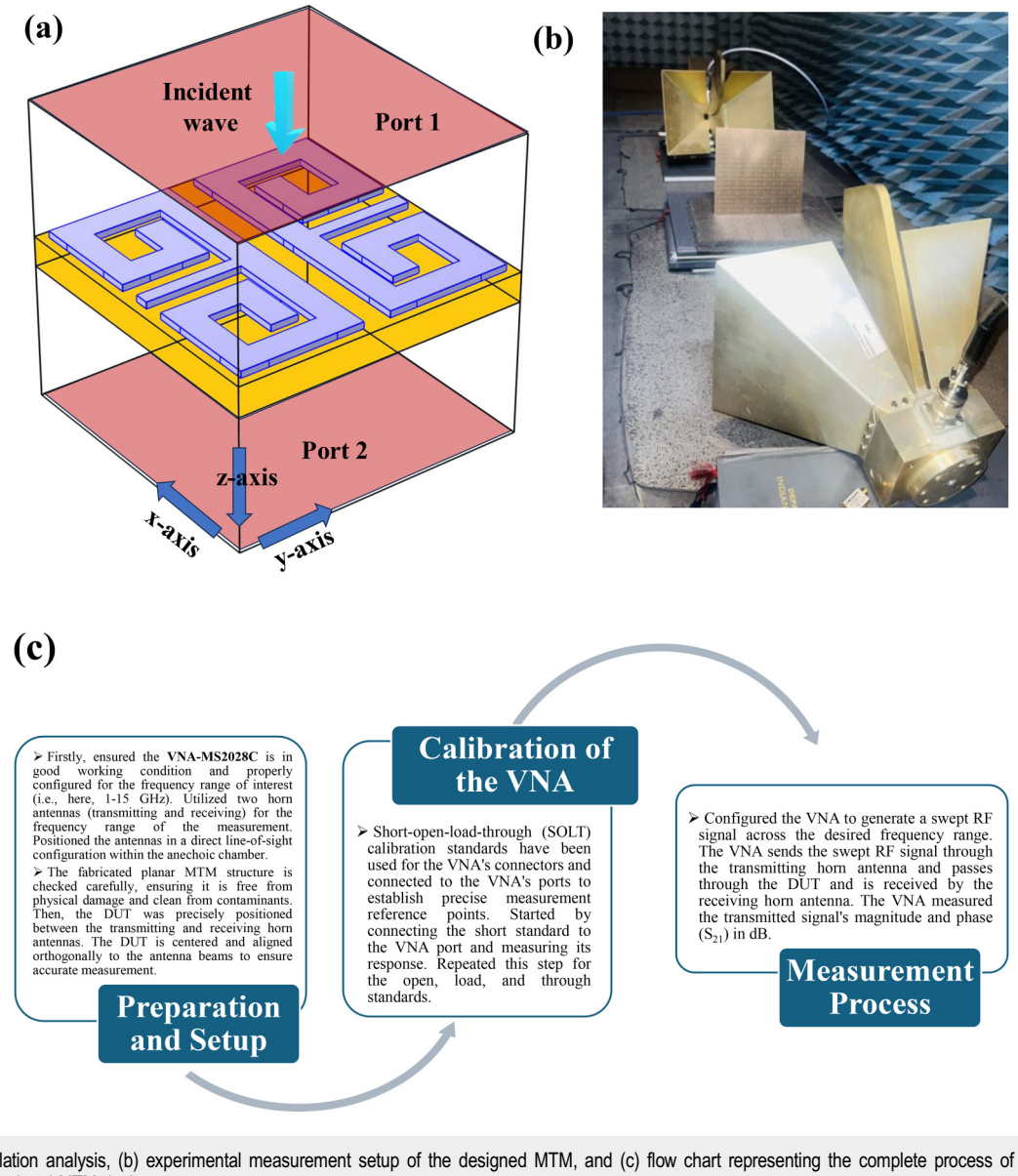


FIG. 2. (a) Simulation analysis, (b) experimental measurement setup of the designed MTM, and (c) flow chart representing the complete process of the measurement process of the simulated MTM design.

18 July 2024 04:56:20

© Author(s) 2024

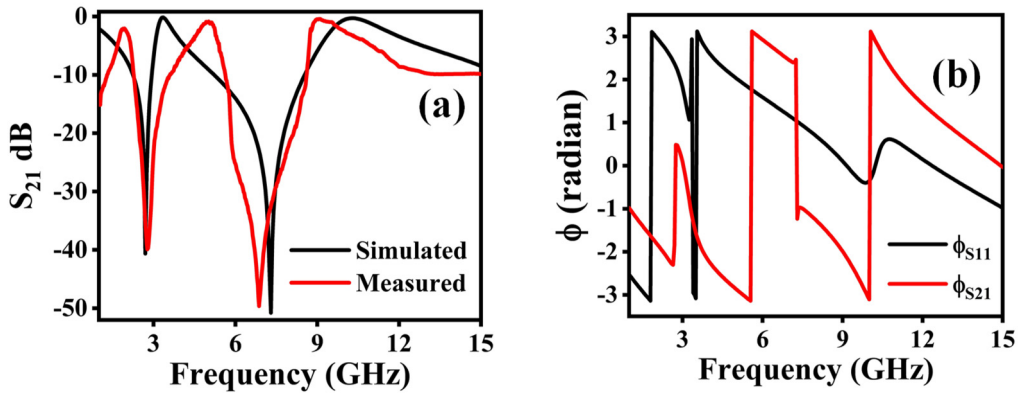


FIG. 3. (a) Transmission coefficient (S_{21}) for the simulated and measured response and (b) the phase plot of the scattering parameter at different incident wave frequencies of the unit cell structure.

test (DUT), a fabricated planar frequency selective surface (FSS), is positioned within the test setup. The VNA determines the transmitted signal (S_{21}) through the DUT. One port of the VNA connects to the transmitting horn antenna, while the other end is attached to the receiver horn antenna, with the DUT positioned between them. The entire antenna setup is placed in an anechoic chamber to minimize external interference and reflections. Calibration of the VNA is vital for obtaining precise measurements and involves the use of short-open-load-through (SOLT) calibration standards connected to the VNA's ports to establish accurate reference points. The VNA generates a swept RF signal transmitted through the DUT, and the transmitted signal is measured by the VNA in terms of magnitude and phase (expressed in dB). Detailed measurement processes are illustrated in a flow chart, as depicted in Fig. 2(c). Figure 3(a) depicts a comparative representation between the simulated and experimentally observed responses of the transmission coefficient (S_{21}) for the fabricated designed structure. Furthermore, the phase plot of the scattering parameters is shown in Fig. 3(b).

The experimental observations align well with the simulation results within the limits of the experimental setup. As depicted in Fig. 3(a), there is a very small and minor mismatch between the experimentally measured and simulated transmission coefficient (S_{21}) outcomes. This mismatch may be attributed due to factors like manufacturing variations, signal interference, reflection noise, and other similar effects. The experimental setup has certain limitations, wherein achieving the desired performance depends on the precise alignment of the two horn antennas. To ensure accurate experimental findings, it is crucial to calibrate the network analyzer properly. All these constraints contribute to a slight shift in resonance frequencies, as evident in the measurement results. Nevertheless, the simulated and experimentally measured data cover the S and C frequency bands. The study on the S_{21} parameter of the designed MTM unit cell indicates the presence of two resonance frequencies at specific points, namely, 2.7 and 7.3 GHz. These resonance frequencies fall within the microwave region, corresponding to the S-band and the C-band, which are commonly used in wireless communication applications. In contrast, Fig. 3(b) illustrates the phase plot of output scattering parameters (S_{11} and S_{21}), demonstrating variations from $-\pi$ to $+\pi$ values with a total phase change of 2π. The bandwidth of S_{21}, which is significant for wireless communication, is shown to be considerable in different frequency bands. For instance, it spans 0.44 GHz (2.42–2.86 GHz) and 1.98 GHz (6.10–8.08 GHz) for the S and C bands, respectively. Importantly, the proposed MTM unit cell has a relatively wider bandwidth for the C band as compared to previously published data,[35–39] as detailed in Table II. This makes it suitable for various wireless communication applications, including satellite and Wi-Fi technologies.[38]

18 July 2024 04:56:20

TABLE II. Comparison of different bandwidths of the proposed work with the previously published work and boldface values denotes the effective bandwidth for C-band.

References	Dimension (physical and electrical)	Resonance frequencies (GHz)	Frequency band	Effective Bandwidth (GHz)
35	8×8 mm^2 ($0.070\lambda \times 0.070\lambda$)	2.61, 6.32, 9.29	S, C, X	0.11, **0.79**, 0.15
36	8×8 mm^2 ($0.06\lambda \times 0.06\lambda$)	2.48, 4.28, 9.36, 13.7	S, C, X, Ku	0.19, **0.4**, 1.3, 0.6
37	10×10 mm^2 ($0.14\lambda \times 0.14\lambda$)	4.20, 10.14, 13.15, 17.1	C, X, Ku	**0.72**, 1.55, 0.17, 0.24
38	10×10 mm^2 ($0.17\lambda \times 0.17\lambda$)	5.0, 6.88, 8.429	C, X	**1.67**, 0.52, 0.98
39	15×15 mm^2 ($0.20\lambda \times 0.20\lambda$)	3.36, 7.41, 10.16	S, C, X	1.51, **0.94**, 0.89
Present work	16×16 mm^2 ($0.07\lambda \times 0.07\lambda$)	2.7 and 7.3	S, C	0.44, **1.98**

© Author(s) 2024

B. Evaluation of effective electromagnetic parameters of the proposed MTM unit cell using the metamaterial extraction method

To observe the frequency-dependent behavior of the metamaterial, researchers have explored several models. Among these models, the Lorentz model is frequently employed for MTM characterization which involves the response of electrons to an electric field (E) by undergoing damped harmonic oscillations, resulting in a polarization field (P). The mathematical representation of the relationship between the polarization field (P) and the electric field (E) can be expressed as follows:[40]

$$\frac{d^2}{dt^2}P_i + \Gamma_L \frac{d}{dt}P_i + \omega_0^2 P_i = \varepsilon_0 \chi_L E_i. \tag{1}$$

In Eq. (1), the terms on the left side correspond to the acceleration of charges, damping coefficient (Γ_L), and the restoring force, which are represented by the first, second, and third terms, respectively. On the right-hand side, the excitation field, also known as the driving field, is present along with the coupling coefficient, χ_L. Employing Eq. (1), researchers assess the electric susceptibility $[\chi_{e,\text{Lorentz}}(\omega)]$, which can be expressed as follows:[40]

$$\chi_{e,\text{Lorentz}}(\omega) = \frac{P_i(\omega)}{\varepsilon_0 E_i(\omega)} = \frac{\chi_L}{-\omega^2 + j\Gamma_L\omega + \omega_0^2}. \tag{2}$$

By utilizing Eq. (2), researchers calculate the electric permittivity ($\varepsilon_{\text{Lorentz}}$)[40] of the Lentz model, which can be expressed as the following equation:

$$\varepsilon_{\text{Lorentz}}(\omega) = \varepsilon_0[1 + \chi_{e,\text{lorentz}}(\omega)]. \tag{3}$$

The Drude model is derived from the Lorentz model of electromagnetic characteristics by eliminating the restoring force from the left side of Eq. (1). Solving this modified equation results in an equation for the electric susceptibility $[\chi_{e,\text{Drude}}(\omega)]$ of the Drude model, which can be represented as follows:[40]

$$\chi_{e,\text{Drude}}(\omega) = \frac{\chi_D}{j\Gamma_D\omega + \omega_0^2}, \tag{4}$$

where Γ_D is the Drude damping coefficient.

Plasma frequency of the Drude model is determined using the following relationship:[40]

$$\chi_D = \omega_p^2. \tag{5}$$

The assessed plasma frequency can serve as an alternative to the coupling coefficient (χ_L) in the Lorentz model. For instance,

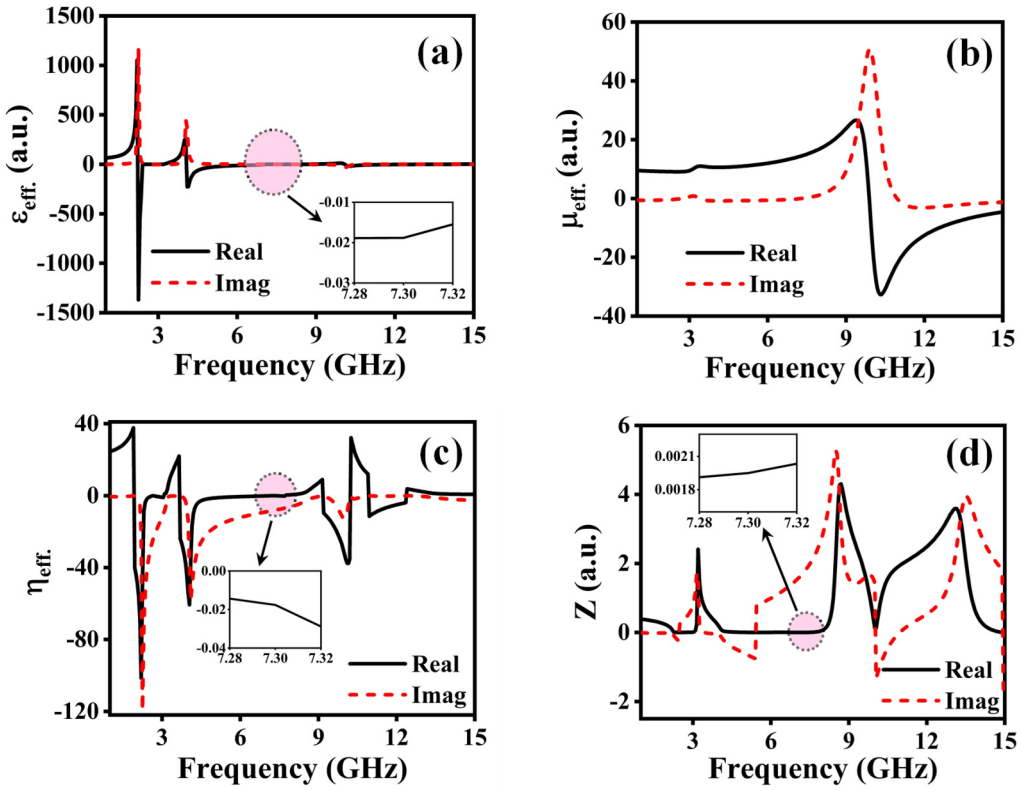


FIG. 4. (a)–(c) illustrate the graphical representations of real and imaginary $\mu_{\text{eff.}}$, $\varepsilon_{\text{eff.}}$, and $\eta_{\text{eff.}}$ for the proposed metamaterial unit cell structure.

18 July 2024 04:56:20

© Author(s) 2024

TABLE III. Effective parameter property.

Parameters	Resonance Frequency (GHz)	Magnitude	Extracted property
Effective permittivity ($\varepsilon_{\text{eff.}}$)	2.7, 7.3	−0.085, −0.588	$\varepsilon_{\text{eff.}} < 0$ (ENG), $\varepsilon_{\text{eff.}} \sim 0$ (ENZ)
Effective permeability ($\mu_{\text{eff.}}$)	2.7, 7.3	9.118, 14.493	$\mu_{\text{eff.}} > 0$ (SNG)
Refractive index ($\eta_{\text{eff.}}$)	2.7, 7.3	−0.144, −0.017	$\eta_{\text{eff.}} < 0$ (LHM), $\eta_{\text{eff.}} \sim 0$ (NZI)
Impedance (Z)	2.7, 7.3	0.0030, 0.0019	$Z \sim 0$

when the coupling coefficient is positive, the Lorentz model exhibits a narrow band of negative real permittivity at frequencies beyond the resonance frequencies. On the other hand, the Drude model shows a wide spectral negative permittivity for frequencies below $\omega < \sqrt{\omega_p^2 - \Gamma_D^2}$. To analyze the metamaterial's scattering parameters, researchers utilized the Drude–Lorentz variables as parameters in numerical simulations. Ziolkowski *et al.* proposed a mathematical method to extract effective parameters such as $\varepsilon_{\text{eff.}}$, $\mu_{\text{eff.}}$, and $n_{\text{eff.}}$ using the NRW method, which is a widely used approach for effective parameter extraction. The NRW method employs transmission coefficient (S_{21}) and reflection coefficient (S_{11}) to derive the effective electromagnet parameters. The simplified equations used in this process are as follows:[41–44]

$$\varepsilon_{\text{eff.}} = \frac{2}{jk_0 d} \times \frac{(1 - V_1)}{(1 + V_1)} = \frac{c}{j\pi v d} \frac{(1 - S_{21} - S_{11})}{(1 + S_{21} + S_{11})}, \tag{6}$$

$$\mu_{\text{eff.}} = \frac{2}{jk_0 d} \times \frac{(1 - V_2)}{(1 + V_2)} = \frac{c}{j\pi v d} \frac{(1 - S_{21} + S_{11})}{(1 + S_{21} - S_{11})}, \tag{7}$$

$$n_{\text{eff.}} = \sqrt{\varepsilon_{\text{eff.}} \times \mu_{\text{eff.}}} = \frac{c}{j\pi v d} \sqrt{\left\{ \frac{(S_{21} - 1)^2 - (S_{11})^2}{(S_{21} + 1)^2 - (S_{11})^2} \right\}}. \tag{8}$$

In the given context, the parameters are defined as follows:

$$V_1 = |S_{21}| + |S_{11}|, \tag{9}$$

$$V_2 = |S_{21}| - |S_{11}|, \tag{10}$$

$$k_0 = \frac{2\pi v}{c}. \tag{11}$$

Here, d represents the metamaterial slab thickness and c represents the velocity of light.

The normalized impedance (Z) can be computed using the following formula:[45,46]

$$Z = \sqrt{\frac{(1 + S_{11})^2 - S_{21}^2}{(1 - S_{11})^2 - S_{21}^2}}. \tag{12}$$

According to the data presented in Figs. 4(a) and 4(b), it is evident that the effective permittivity ($\varepsilon_{\text{eff.}}$) of the proposed MTM structure is negative at resonance frequencies of 2.7 and 7.3 GHz, while it approaches near to zero (ENZ) at these frequencies as well [shown in the inset of Fig. 4(a)]. On the other hand, the effective permeability ($\mu_{\text{eff.}}$) remains positive across all resonance frequencies for the designed metamaterial structure, as shown in Fig. 4(b). Furthermore, Fig. 4(c) shows that the effective refractive index ($\eta_{\text{eff.}}$) exhibits negative and near-zero values at dual resonance frequencies, which correspond to the S and C bands [shown in the inset of Fig. 4(c)]. In Fig. 4(d), the normalized impedance (Z) demonstrates positive and near-zero real values across the resonance frequencies [represented in the inset of Fig. 4(d)]. Smaller values of normalized impedance at resonance frequencies indicate the passive nature (i.e., which have fixed properties determined by their structural design) of the designed metamaterial unit cell. Table III provides a summary of the overall parameter properties.

EMR plays a crucial role in metamaterial design and development as it contributes to the compactness of the structure. Effective medium theory (EMT) is employed in designing the metamaterial structure, and according to EMT, the unit cell dimension should be smaller than the operating wavelength.[47] In the case of the proposed structure, it is observed that the unit cell size is 16 mm, which is significantly smaller than the operating wavelength for all quad-frequency bands. As a result, the proposed structure meets the EMT condition, leading to the desired metamaterial (MTM) characteristics. The formula provided below can be used to calculate EMR,[47]

$$\text{EMR} = \frac{\text{Operating Wavelength for the unit cell}}{\text{length of the unit cell}}. \tag{13}$$

The proposed MTM unit cell exhibits a relatively high EMR and compact dimensions, enabling it to cover all dual frequency bands (S and C) commonly used in satellite and multi-band applications. The calculated EMR value for the designed metamaterial at its operating resonance frequency of 2.7 GHz is 6.93. The relatively high EMR value contributes to the improved compactness of the structure, making it a promising choice for various practical applications.

C. Optimization of the designed MTM geometry

The optimization process for the proposed MTM structure involves five design steps, as depicted in Fig. 5. Design 1 consists of four L-shaped metallic strips placed at the corners of the substrate, resulting in resonance frequencies at 11.65 GHz for the transmission scattering parameter (S_{21}), as shown in Fig. 6. Moving on to design 2, four horizontal metallic wires are added to the middle edges, leading to resonance frequencies at 7.6 GHz. Design 3 involves the addition of a central metallic strip incorporated into

18 July 2024 04:56:20

© Author(s) 2024

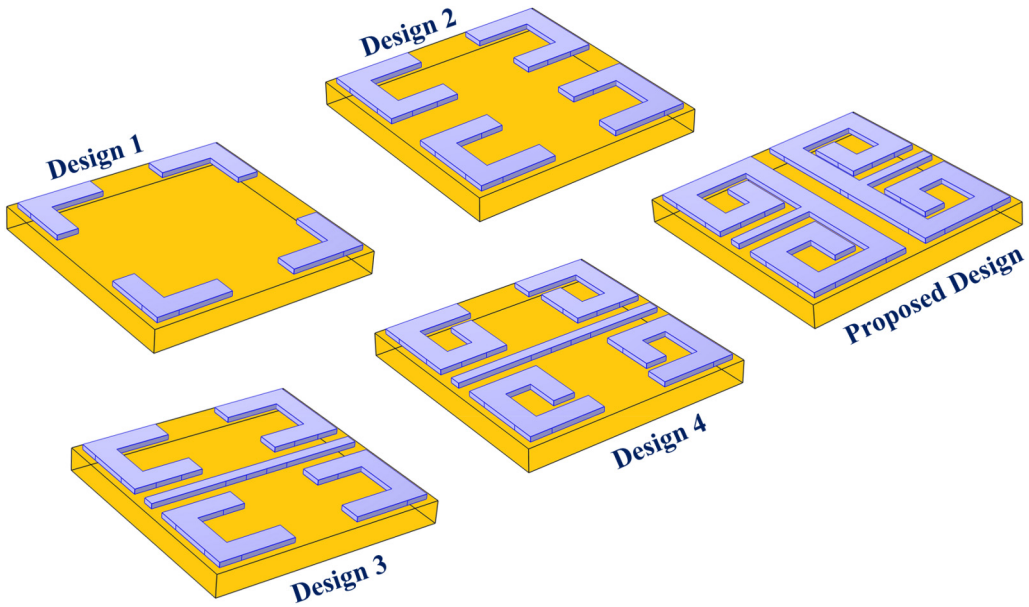


FIG. 5. The methodological approach employed in creating the MTM structure.

design 2, generating resonance frequencies at 7.45 GHz. All these designs resonate in single S bands.

Design 4 builds upon design 3 by adding four more vertical metallic strips, resulting in resonance frequencies at 6.25 and 14.6 GHz. But this design results in narrow band resonance frequencies although dual-band resonance occurs. Finally, in design 5, two vertical strips are joined at their ends in the center of design 4, forming an MTM with a mirror image shape. This design reflects dual-band operating resonance frequencies at 2.7 and 7.3 GHz in the S and C bands, respectively, within the microwave region, as depicted in Fig. 6. The resonance frequencies of the designed structure are due to the formation of a capacitor and an inductor for the gap region and metallic strips, respectively, which becomes an LC circuit like an LCR tank circuit.[48] The resonance frequency (f) can be found using the formula:[49]

$$f = \frac{1}{2\pi\sqrt{LC}}. \tag{14}$$

D. Electromagnetic field distribution of the designed metamaterial

The presence of induced current in the metallic component of metamaterials (MTMs) leads to the generation of a scattering field when a plane wave interacts with an MTM surface. Maxwell's equations provide a comprehensive explanation for how the electric field (E), surface current (J), and magnetic field (B) interact within an MTM. When a current flows through MTM conducting elements, it generates a magnetic field. However, any change in the magnetic field can induce an electromotive force. Consequently, Maxwell's formulation of Ampere's and Faraday's laws yields a set of four differential equations that establish connections between electric and magnetic fields, elucidating these phenomena.[50] The behavior of electromagnetic waves in a medium is influenced by several key factors, including permeability (μ), permittivity (ε), and

18 July 2024 04:56:20

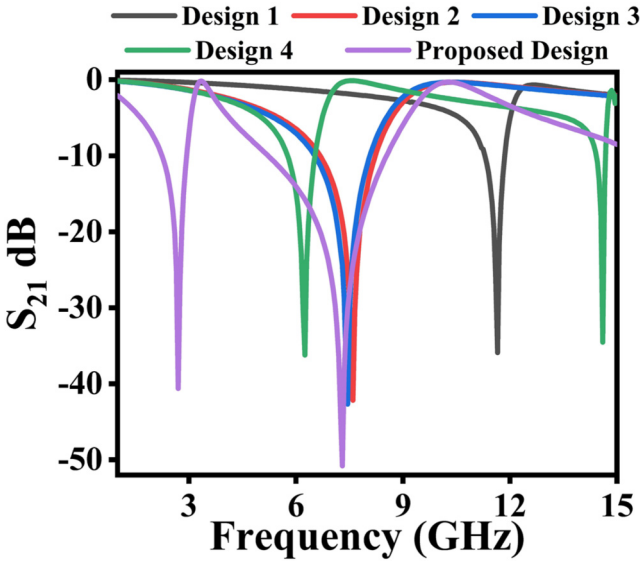


FIG. 6. The variation of the scattering parameter S_{21} (transmission coefficient) concerning the frequency of the incident electromagnetic plane wave, observed across different MTM structures.

© Author(s) 2024

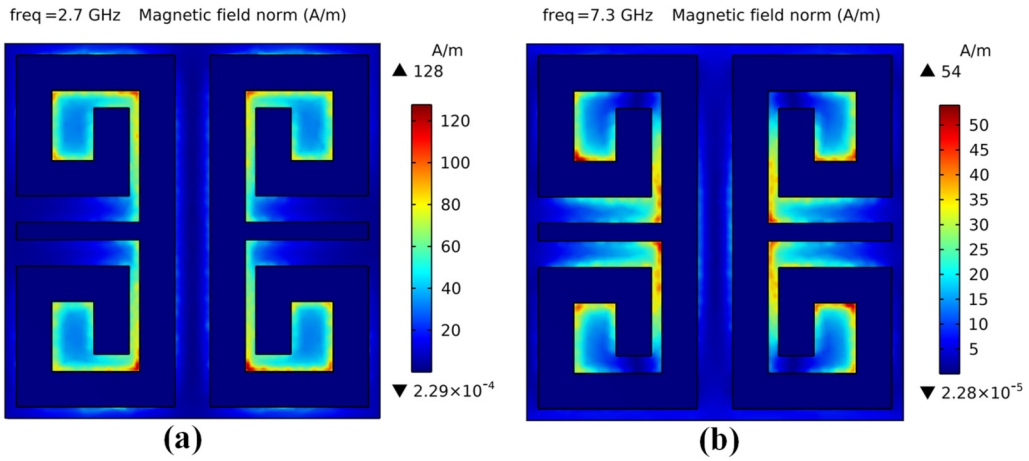


FIG. 7. Magnetic field distribution corresponding to resonance frequencies at (a) 2.7 GHz and (b) at 7.3 GHz.

conductivity (σ) of that medium. These material properties, along with the boundary conditions of the medium, play a significant role in determining the electromagnetic characteristics.

Magnetic field distributions for two resonance frequencies, i.e., 2.7 and 7.3 GHz are shown in Figs. 7(a) and 7(b), respectively. Magnetic field and surface current density are correlated, as per Maxwell's equations. In Fig. 7(a), high magnetic field intensities are shown on the inner side of the structure, near the edges where a larger current density is observed. In Fig. 7(b), a magnetic field with low intensity is depicted, attributed to the low current intensity at this frequency. However, the field intensity remains low in other regions of the MTM structure.

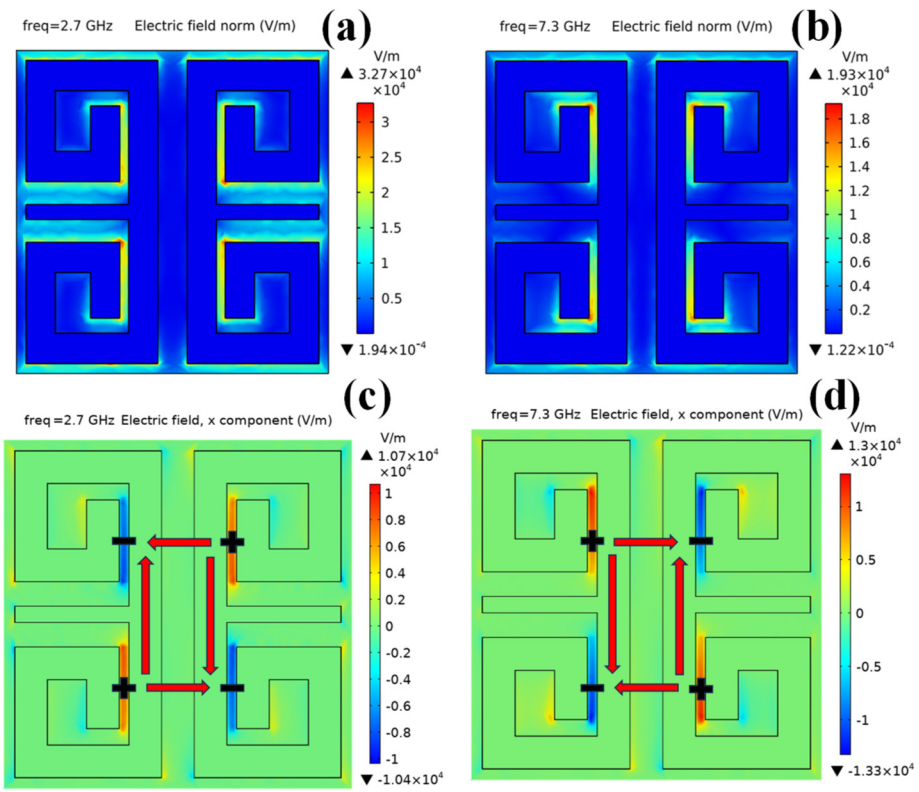


FIG. 8. Electric field distribution corresponding to resonance frequencies at (a) 2.7 GHz and (b) at 7.3 GHz. Electric field *x*-component showing electric quadrupole mode resonance at (c) 2.7 GHz and (d) 7.3 GHz.

18 July 2024 04:56:20

© Author(s) 2024

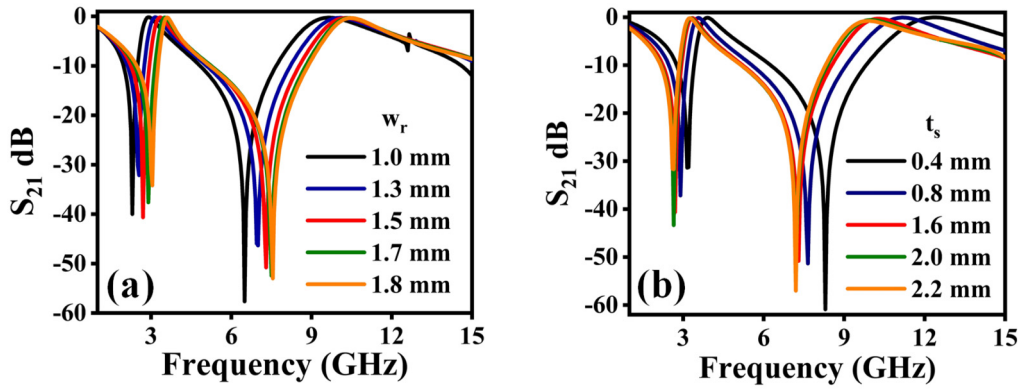


FIG. 9. The response of the scattering parameter (S_{21}) relating to (a) different resonator widths (w_r) and (b) different substrate thicknesses (t_s).

Figures 8(a) and 8(b) display the electric field distribution. It is evident that changes in the magnetic field affect the induced electric field. Intense electric fields are observed in areas where magnetic fields experience the most rapid rate of change, as illustrated by the comparison between Figs. 8(a) and 7(a). Due to the capacitive effect, a strong electric field is detected near the gap in the MTM structure. As a result, at 2.7 GHz, the outer and middle horizontal edges exhibit a stronger electric field than the remaining area. On the other end, in Fig. 8(b), at 7.3 GHz, a low electric field intensity is observed due to low electric current resulting in less magnetic field intensity. Furthermore, the electric field is neutralized when the magnetic field is constant in the resonator. Due to the interconnected nature of electromagnetic fields and currents, both play a vital role in the resonances at the frequencies mentioned above. In Figs. 8(c) and 8(d), the field distribution of the x-component of electric field (E_x) is presented which indicates the formation of electric quadrupole mode resonance at the metal–dielectric interface. The proposed structure's resonance is primarily caused by the excitation of surface plasmon polaritons (SPPs) due to electric quadrupole mode resonance, as indicated by the field distribution, which also shows the accumulation of positive and negative charges in the gap region of the resonator structure.[13,51–53]

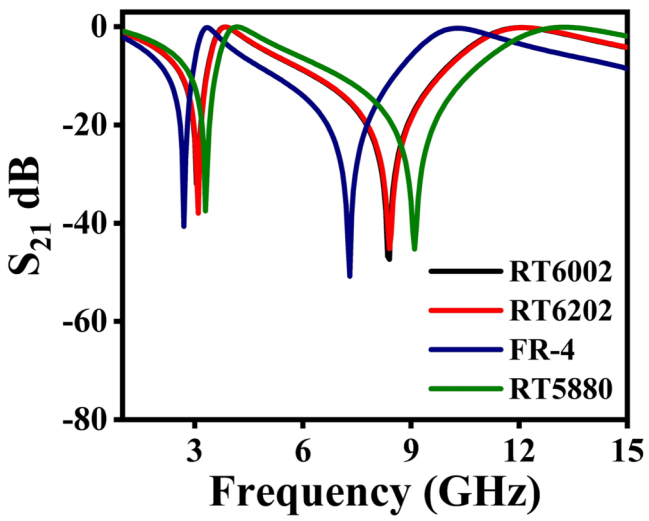


FIG. 10. The variation of the scattering parameter (S_{21}) with different materials as substrates.

E. Simulation-based parametric analysis of the proposed MTM unit cell structure

1. Effect of resonator width (w_r) on resonance frequency

Figure 9(a) demonstrates the relationship between the scattering parameter (S_{21}) and the resonator width (w_r). Altering the width has an impact on inductance, consequently leading to changes in the resonance frequency of the proposed MTM structure.[37,38] Notably, a significant blue shift is observed in the S_{21} parameter when w_r is increased from 1 to 1.8 mm leading to a rise in inductance, indicating an increase in the resonance frequency. The observed shift in resonance frequencies also imparts tunability characteristics of the designed MTM structure.

2. The impact of varying the thickness (t_s) of the FR-4 dielectric substrate

Substrate FR-4 thickness (t_s) is varied from 0.4 to 2.2 mm to analyze its influence on the scattering parameter (S_{21}), as shown in Fig. 9(b). Substrate thickness has less impact on resonance frequency, although this change is less observed at the lower-valued resonance frequency (in the S-band). However, at high values, particularly in the C band, the resonance frequency shift becomes more noticeable with increasing thickness of the FR-4 substrate. This is because the substrate functions as a dielectric medium, leading to the creation of capacitance between the two waveguide ports. As the substrate thickness increases, parallel capacitance also increases, causing a decrease in the resonance frequency.[38] Figure 9(b) illustrates that the MTM

18 July 2024 04:56:20

© Author(s) 2024

TABLE IV. Various properties and S21 parameter results of different substrate materials.

S. No.	Substrate	Electric permittivity	Tangent loss	Resonance frequency of S_{21} (GHZ)	Frequency band	Magnitude of S_{21} (dB)
1.	Roger's RT6002	2.94	0.0012	3.1 and 8.4	S and X	−31.9 and −47.3
2.	Roger's RT6202	2.90	0.0015	3.1 and 8.4	S and X	−37.9 and −45.1
3.	Roger's RT5880	2.20	0.0009	3.3 and 9.1	S and X	−37.4 and −45.2
4.	FR−4	4.30	0.025	2.7 and 7.3	S and C	−40.6 and −50.7

unit cell achieves the required resonance frequencies in both S and C bands when the substrate thickness is set at 1.6 mm. Therefore, this variation of w_r and t_s shows the tunability characteristics of the designed structure with a shift in the resonance frequency.

3. Effect of different dielectric substrate materials

This study examines how changing the dielectric substrate material affects the performance of MTMs. When using Roger's versions instead of the FR-4 substrate, the frequency response of S21 shifts toward higher frequencies. This shift indicates that the choice of substrates significantly influences MTM performance, as depicted in Fig. 10. Each Roger's version demonstrates dual resonance frequencies in S and X bands with varying magnitudes, but not in C-bands. In contrast, the FR-4 substrate exhibits dual resonances at 2.7 and 7.3 GHz in S and C bands, with magnitudes of −40.6 and −50.7 dB, respectively. The discussion clarifies the significance of the FR-4 dielectric material as a suitable substrate for the proposed MTM structure. These above findings are summarized in Table IV.

F. MTM as a refractive index sensor

The transmission coefficient (S_{21}) change with respect to various refractive index (RI) values has been examined in the following section to evaluate RI sensitivity. The study of RI sensing properties involves altering the surrounding environment of the MTM to correspond with changes in its refractive index, consequently inducing a shift in the resonance frequency of the proposed MTM, as shown in Fig. 11(b). Sensitivity (S_R) serves as a crucial benchmark for evaluating sensing performance and is defined as follows:[27]

$$S_R = \frac{\Delta f}{\Delta n}. \tag{15}$$

Transmission coefficient spectra show a red shift as the refractive index of the surrounding environment changes from 1.00 to 1.20, as shown in Fig. 11(a). Figure 11(b) illustrates the correlation between the shift in resonance frequency shift (Δf) and the variations in refractive index (Δn) of the surrounding environment. Additionally, the resonance frequency of the proposed MTM exhibits a linear decrease with the rise in the refractive index of the surrounding environment. Moreover, it is clarified that sensitivities S_{R1} and S_{R2}, corresponding to the first and second resonance peaks in the S and C bands, are about 1 and 3 GHz/RIU, respectively. These sensitivities surpass those of the previously reported MTM sensors, as reported in Table V. Thus, it can be said that in the GHz band, the suggested MTM sensor has a high refractive index sensitivity.

G. Practical applications and commercialization of MTM sensors

The designed MTM structure is highly suitable for chemical and biological sensing applications, offering high sensitivity of 1 and 3 GHz/RIU at dual resonance frequencies. The MTM

18 July 2024 04:56:20

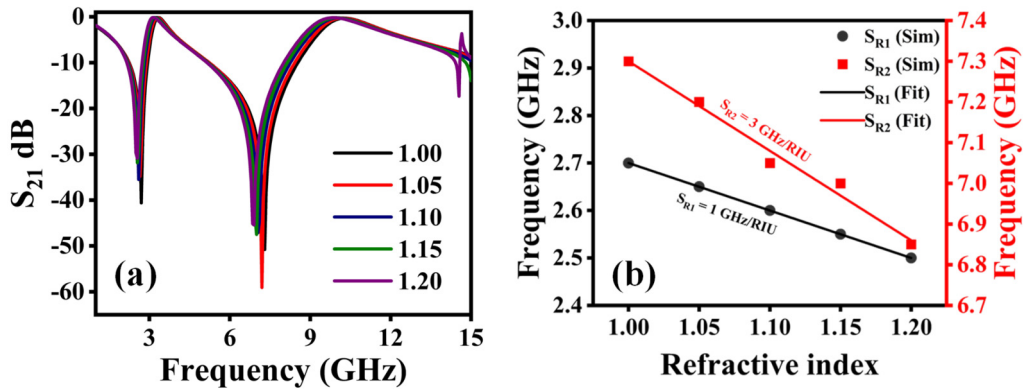


FIG. 11. (a) Transmission coefficient (S_{21}) spectra of the proposed MTM sensor using different RI values of the surrounding environment. (b) The variation of simulated resonance frequency (square and circle symbols) and linear fitting (solid line) with different RI values of the surrounding environment.

© Author(s) 2024

TABLE V. Comparison of the sensitivity of the proposed MTM structure with the previous literature work.

Reference	Designed structure	Frequency band	Sensitivity
54	SRR	Single band (C band)	$S_R = 1.69$ GHz/RIU
27	Frequency selective surface (FSS)	Single band (X band)	$S_R = 1.31$ GHz/RIU
28	Periodical circle rings resonator	Single band (C band)	$S_R = 0.3537$ GHz/RIU
29	Microwave resonator	Dual band (C and X bands)	$S_{R_1} = 1.116$ GHz/RIU $S_{R_2} = 2.357$ GHz/RIU
55	Corona-shaped metamaterial resonator	Single band (S band)	$S_R = 0.1825$ GHz/RIU
This work	Metamaterial structure	Dual band (S and C bands)	$S_{R_1} = 1$ GHz/RIU $S_{R_2} = 3$ GHz/RIU

structure's dual-band operation within the GHz frequency range makes it a promising candidate for advanced optoelectronic devices. Its near-zero index and left-handed behavior are particularly advantageous for non-destructive testing and evaluation methods,[56] which are critical for material characterization and quality control in manufacturing industries.

This study has focused on developing technology that is both easy to fabricate and straightforward to implement within the GHz frequency range. This focus on simplicity and ease of production ensures that this technology can be readily adopted and scaled up for mass production. The level of accuracy achieved in the design process allows the technology to be classified within the microwave region of the electromagnetic spectrum, which spans from 1 to 15 GHz frequencies. This classification is important because a variety of applications, such as radar systems satellite and wireless telecommunications, depend on technology that operates in the microwave region. These technologies are very fascinating for commercial and industrial use due to their high-precision production capabilities and ease of fabrication. This easy manufacturing procedure lowers expenses and increases mass productivity, allowing for the large-scale production of these technologies. Therefore, this study's emphasis on easy-to-fabricate and implement technologies in the GHz region positions it well within the current technological landscape. This compatibility with mass production systems further boosts its potential for widespread adoption and integration into various applications that rely on microwave technology.

IV. CONCLUSION

The current study presents a novel design, fabrication, and simulation of an MTM structure capable of dual-band operation with a wide bandwidth in the C band and refractive index sensing application. The FEM-based high-frequency electromagnetic simulator in the frequency domain is used to design, simulate, and extract the MTM's scattering parameter. The experimental measurements were compared to the simulated results and found to be well-matched. The resonance frequencies showed minimal frequency shift, indicating that any differences between the simulated and measured values fell within the acceptable range. This study investigated the effective medium parameters of the designed MTM structure within the frequency range of 1–15 GHz using the NRW method. The analysis of the proposed mirror-shaped metamaterial revealed the presence of effective negative and near-zero refractive index (NZI) and effective negative and near-zero epsilon (ENZ) for dual-band operations. The electromagnetic field distributions were also plotted at the resonance frequencies. The EMR of the designed structure was measured to be 6.93, which confirms the compactness of the metamaterial and its applicability across all dual-frequency bands in the microwave region. Utilizing the suggested MTM as a refractive index sensor has the potential to attain significant sensitivity, specifically 1 GHz/RIU and 3 GHz/RIU. This research introduces a potential alternative approach for creating advanced multi-function sensors with exceptional sensing capabilities, making them suitable for applications in medical diagnostics, detection, and biochemical sensing.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the initiatives and support from the TIFAC—Centre of Relevance and Excellence in Fiber Optics and Optical Communication at Delhi College of Engineering, now DTU, Delhi, under the Mission REACH program of Technology Vision-2020, Government of India. One of the authors, Mr. Ankit, would like to acknowledge (i) the Central Research Facility (CRF) of IIT-Delhi for facilitating with measurement of fabricated design using RF and Antenna Measurement setup and (ii) Delhi Technological University, Delhi, for support in terms of Research Fellowship.

18 July 2024 04:56:20

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Ankit: Investigation (lead); Methodology (lead); Resources (equal); Software (lead); Validation (equal); Visualization (equal); Writing – original draft (lead). **Monu Nath Baitha:** Formal analysis (supporting); Supervision (equal); Writing – review & editing (equal). **Kamal Kishor:** Formal analysis (equal); Resources (equal); Supervision (equal); Validation (equal); Writing – review & editing (equal). **Ravindra Kumar Sinha:** Resources (equal); Supervision (lead); Writing – review & editing (equal).

© Author(s) 2024

DATA AVAILABILITY

The data that support the findings of this study are available within the article.

REFERENCES

[1]N. Krishna and K. G. Padmasine, "A review on microwave band pass filters: Materials and design optimization techniques for wireless communication systems," Mater.: Sci. Semicond. Process. **154**, 107181 (2023).
[2]V. Kumar and D. K. Gupta, *Radar Remote Sens* (Elsevier, 2022), pp. 29–48.
[3]D. K. Gupta, S. Prashar, S. Singh, P. K. Srivastava, and R. Prasad, *Radar Remote Sens* (Elsevier, 2022), pp. 3–27.
[4]D. Shin, Y. Urzhumov, Y. Jung, G. Kang, S. Baek, M. Choi, H. Park, K. Kim, and D. R. Smith, "Broadband electromagnetic cloaking with smart metamaterials," Nat. Commun. **3**(1), 1213 (2012).
[5]D. Shin, J. Kim, C. Kim, K. Bae, S. Baek, G. Kang, Y. Urzhumov, D. R. Smith, and K. Kim, "Scalable variable-index elasto-optic metamaterials for macroscopic optical components and devices," Nat. Commun. **8**(1), 16090 (2017).
[6]J. B. Pendry, D. Schurig, and D. R. Smith, "Controlling electromagnetic fields," Science **312**(5781), 1780–1782 (2006).
[7]R. A. Shelby, D. R. Smith, and S. Schultz, "Experimental verification of a negative index of refraction," Science **292**(5514), 77–79 (2001).
[8]N. Shankhwar and R. K. Sinha, *Zero Index Metamaterials* (Springer Singapore, Singapore, 2021), pp. 57–82.
[9]Ankit, K. Kishor, and R. K. Sinha, "Design, fabrication, and characterization of epsilon negative and near-zero index metasurface," Appl. Phys. A **130**(2), 98 (2024).
[10]Ankit, K. Kishor, and R. K. Sinha, "SOI based metasurface for broadband perfect reflection in visible spectrum," J. Opt. **26**(4), 045101 (2024).
[11]J. Zhang, S. Yan, and G. A. E. Vandenbosch, "Metamaterial-inspired dual-band frequency-reconfigurable antenna with pattern diversity," Electron. Lett. **55**(10), 573–574 (2019).
[12]C.-X. Mao, S. Gao, Y. Wang, Q.-X. Chu, and X.-X. Yang, "Dual-band circularly polarized shared-aperture array for C-/X-band satellite communications," IEEE Trans. Antennas Propag. **65**(10), 5171–5178 (2017).
[13]P. Agarwal, K. Kishor, and R. K. Sinha, "Ultrasensitive dual-band terahertz metasurface sensor based on all InSb resonator," Opt. Commun. **522**, 128667 (2022).
[14]T. Ergin, N. Stenger, P. Brenner, J. B. Pendry, and M. Wegener, "Three-dimensional invisibility cloak at optical wavelengths," Science **328**(5976), 337–339 (2010).
[15]X. Ankit, K. Kishor, and R. K. Sinha, "Design and analysis of wavelength tunable metamaterial reflector," Results Opt. **10**, 100366 (2023).
[16]X. Hengbo, "Design, simulation, and measurement of a multiband tunable metamaterial filter," Opt. Mater. **127**, 112253 (2022).
[17]D. Shin, Y. Urzhumov, D. Lim, K. Kim, and D. R. Smith, "A versatile smart transformation optics device with auxetic elasto-electromagnetic metamaterials," Sci. Rep. **4**(1), 4084 (2014).
[18]J. Lv, M. Zhou, Q. Gu, X. Jiang, Y. Ying, and G. Si, "Metamaterial lensing devices," Molecules **24**(13), 2460 (2019).
[19]K. Lee, Y. Jung, G. Kang, H. Park, and K. Kim, "Active phase control of a Ag near-field superlens via the index mismatch approach," Appl. Phys. Lett. **94**(10), 101113 (2009).
[20]S. Banerjee, C. S. Amith, D. Kumar, G. Damarla, A. K. Chaudhary, S. Goel, B. P. Pal, and D. Roy Chowdhury, "Ultra-thin subwavelength film sensing through the excitation of dark modes in THz metasurfaces," Opt. Commun. **453**, 124366 (2019).
[21]N. Navaratna, Y. J. Tan, A. Kumar, M. Gupta, and R. Singh, "On-chip topological THz biosensors," Appl. Phys. Lett. **123**(3), 033705 (2023).
[22]M. Gupta and R. Singh, "Terahertz sensing with optimized Q/Veff metasurface cavities," Adv. Opt. Mater. **8**(16), 1–7 (2020).
[23]K. V. Sreekanth, Q. Ouyang, S. Sreejith, S. Zeng, W. Lishu, E. Ilker, W. Dong, M. ElKabbash, Y. Ting, C. T. Lim, M. Hinczewski, G. Strangi, K. T. Yong, R. E. Simpson, and R. Singh, "Phase-change-material-based low-loss visible-frequency hyperbolic metamaterials for ultrasensitive label-free biosensing," Adv. Opt. Mater. **7**(12), 1–9 (2019).
[24]M. Chen, L. Singh, N. Xu, R. Singh, W. Zhang, and L. Xie, "Terahertz sensing of highly absorptive water-methanol mixtures with multiple resonances in metamaterials," Opt.: Express **25**(13), 14089 (2017).
[25]J. Xu, D. Liao, M. Gupta, Y. Zhu, S. Zhuang, R. Singh, and L. Chen, "Terahertz microfluidic sensing with dual-torus toroidal metasurfaces," Adv. Opt. Mater. **9**(15), 1–8 (2021).
[26]W. X. Lim, M. Manjappa, P. Pitchappa, and R. Singh, "Shaping high-Q planar fano resonant metamaterials toward futuristic technologies," Adv. Opt. Mater. **6**(19), 1800502 (2018).
[27]W. Zhang, J. Y. Li, J. Xie, and R. Xu, "High sensitivity refractive index sensor based on frequency selective surfaces absorber," IEEE Sensors Lett. **2**(3), 107–115 (2018).
[28]W. Zhang, J.-Y. Li, and J. Xie, "High sensitivity refractive index sensor based on metamaterial absorber," Prog.: Electromagn. Res. M **71**, 107–115 (2018).
[29]S. Hu and D. Liu, "Dual-band microwave resonators based on cross hollow structures for refractive index sensing," Phys. Scr. **95**(8), 085504 (2020).
[30]K. Kishor, M. N. Baitha, R. K. Sinha, and B. Lahiri, "Tunable negative refractive index metamaterial from V-shaped SRR structure: Fabrication and characterization," J. Opt. Soc. Am. B **31**(7), 1410 (2014).
[31]T. M. Hossain, M. F. Jamlos, M. A. Jamlos, P. J. Soh, M. I. Islam, and R. Khan, "Modified H-shaped DNG metamaterial for multiband microwave application," Appl. Phys. A Mater. Sci. Process. **124**(2), 1–7 (2018).
[32]I. N. Idrus, M. R. I. Faruque, S. Abdullah, M. U. Khandaker, N. Tamam, and A. Sulieman, "An oval-square shaped split ring resonator based left-handed metamaterial for satellite communications and radar applications," Micromachines **13**(4), 578 (2022).
[33]A. Dhouibi, S. N. Burokur, A. de Lustrac, and A. Priou, "Study and analysis of an electric Z-shaped meta-atom," Adv. Electromagn. **1**(2), 64–70 (2012).
[34]K. Kishor, M. N. Baitha, and R. K. Sinha, "Design and simulation of 'i' shaped split ring resonator metamaterial at optical communication window around 1.55 μm," Optik **126**(23), 4708–4711 (2015).
[35]M. Rashedul Islam, M. Tariqul Islam, M. Moniruzzaman, M. Samsuzzaman, B. Bais, H. Arshad, and G. Muhammad, "Square enclosed circle split ring resonator enabled epsilon negative (ENG) near zero index (NZI) metamaterial for gain enhancement of multiband satellite and radar antenna applications," Results Phys. **19**, 103556 (2020).
[36]M. T. Islam, M. Moniruzzaman, T. Alam, M. Samsuzzaman, Q. A. Razouqi, and A. F. Almutairi, "Realization of frequency hopping characteristics of an epsilon negative metamaterial with high effective medium ratio for multiband microwave applications," Sci. Rep. **11**(1), 16898 (2021).
[37]M. Moniruzzaman, M. T. Islam, I. Hossain, M. S. Soliman, M. Samsuzzaman, and S. H. A. Almalki, "Symmetric resonator based tunable epsilon negative near zero index metamaterial with high effective medium ratio for multiband wireless applications," Sci. Rep. **11**(1), 21842 (2021).
[38]M. Bellal Hossain, M. Rashed Iqbal Faruque, A. S. Alshammari, and M. Tariqul Islam, "Wide bandwidth enriched symmetric hexagonal split ring resonator based triple band negative permittivity metamaterial for satellite and Wi-Fi applications," Results Phys. **37**, 105511 (2022).
[39]I. Hossain, M. Tariqul Islam, N. Binti Mohd Sahar, A. Mohamed Moubark, M. S. Soliman, and M. Samsuzzaman, "Tunable cross-coupled metallic elements loaded epsilon negative and near-zero index characteristics based metamaterial," J. Magn. Magn. Mater. **565**, 170226 (2023).
[40]N. Engheta and R. Ziolkowski, *Electromagnetic Metamaterials: Physics and Engineering Explorations* (John Wiley & Sons, 2006).
[41]E. J. Rothwell, J. L. Frasch, S. M. Ellison, P. Chahal, and R. O. Ouedraogo, "Analysis of the Nicolson-Ross-Weir method for characterizing the electromagnetic properties of engineered materials," Prog. Electromagn. Res. **157**, 31–47 (2016).

18 July 2024 04:56:20

© Author(s) 2024

[42]S. S. Islam, M. R. I. Faruque, and M. T. Islam, "A new direct retrieval method of refractive index for the metamaterial," Curr. Sci. **109**(2), 337–342 (2015).
[43]R. W. Ziolkowski and E. Heyman, "Wave propagation in media having negative permittivity and permeability," Phys.: Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. **64**(5), 15 (2001).
[44]X. Chen, T. M. Grzegorczyk, B. I. Wu, J. Pacheco, and J. A. Kong, "Robust method to retrieve the constitutive effective parameters of metamaterials," Phys.: Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top. **70**(1), 7 (2004).
[45]C. Sabah, "Tunable metamaterial design composed of triangular split ring resonator and wire strip for s- and c- microwave bands," Prog. Electromagn. Res. B **22**, 341–357 (2010).
[46]R. W. Ziolkowski, "Design, fabrication, and testing of double negative metamaterials," IEEE Trans. Antennas Propag. **51**(7), 1516–1529 (2003).
[47]A. F. Almutairi, M. S. Islam, M. Samsuzzaman, M. T. Islam, N. Misran, and M. T. Islam, "A complementary split ring resonator based metamaterial with effective medium ratio for C-band microwave applications," Results Phys. **15**, 102675 (2019).
[48]Y. Guo and J. Zhou, "Dual-band-enhanced transmission through a subwavelength aperture by coupled metamaterial resonators," Sci. Rep. **5**(1), 8144 (2015).
[49]M. Moniruzzaman, M. T. Islam, N. Misran, M. Samsuzzaman, T. Alam, and M. E. H. Chowdhury, "Inductively tuned modified split ring resonator based quad band epsilon negative (ENG) with near zero index (NZI) metamaterial for multiband antenna performance enhancement," Sci. Rep. **11**(1), 11950 (2021).
[50]M. M. Bait-Suwailam, *Electromagnetic Fields Waves* (IntechOpen, 2019).
[51]Z. Wang, R. Zhang, and J. Guo, "Quadrupole mode plasmon resonance enabled subwavelength metal-dielectric grating optical reflection filters," Opt. Express **26**(1), 496 (2018).
[52]H. Liu, K. Wang, J. Gao, M. Liu, H. Zhang, and Y. Zhang, "Dirac semimetal and an all dielectric based tunable ultrasensitive terahertz sensor with multiple bound states in the continuum," Opt. Express **30**(26), 46471 (2022).
[53]M. Aftab, M. S. Mansha, T. Iqbal, and M. Farooq, "Surface plasmon excitation: Theory, configurations, and applications," Plasmonics 1–19 (2023).
[54]Y. Tian, S. Hu, X. Huang, Z. Yu, H. Lin, and H. Yang, "Low-loss planar metamaterials electromagnetically induced transparency for sensitive refractive index sensing," J.: Phys. D. Appl. Phys. **50**(40), 405105 (2017).
[55]N. Dadouche, Z. Mezache, J. Tao, E. Ali, M. Alsharef, A. Alwabli, A. Jaffar, A. Alzahrani, and A. Berazguia, "Design and fabrication of a novel corona-shaped metamaterial biosensor for cancer cell detection," Micromachines **14**(11), 2114 (2023).
[56]T. ul Haq, C. Ruan, X. Zhang, and S. Ullah, "Complementary metamaterial sensor for nondestructive evaluation of dielectric substrates," Sensors **19**(9), 2100 (2019).

18 July 2024 04:56:20

© Author(s) 2024

Journal of Fluorescence
https://doi.org/10.1007/s10895-024-03781-y

RESEARCH

Quantum Computational and Spectroscopic Investigation of 3-aminosalicylic Acid

Shahid Husain[1] · Nupur Pandey[1] · Nisha Fatma[1] · Sanjay Pant[1] · Mohan Singh Mehata[2]

Received: 8 March 2024 / Accepted: 6 June 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Quantum chemical calculations of 3-aminosalicylic acid (3ASA) (monomer and dimer forms) have been performed using DFT and TD-DFT theories with B3LYP/6–311 G(d,p) functional level in the ground and excited states. Using TD-DFT with IEF-PCM model, the electronic spectra of 3ASA in solvents were computed and correlated with the experimental data. The theoretically calculated absorption and emission maxima of 3ASA (monomer) are observed in the range of 343 - 347 nm ($S_0 \rightarrow S_1$ transition) and 429 – 448 nm ($S_1 \rightarrow S_0$ transition), respectively. The natural bond orbital (NBO) analysis shows that charge transfer interaction contributes significantly to stabilize the molecular system. In the case of dimer, hydrogen bonding plays a dominant role in stabilizing the molecular framework. Additionally, the obtained nonlinear optical (NLO) properties: polarizability (13.86×10^{-24} e.s.u for monomer and 29.46×10^{-24} e.s.u. for dimer), first-order hyperpolarizability (4.21×10^{-30} e.s.u for monomer and 0.18×10^{-30} e.s.u for dimer) and second-order hyperpolarizability (7.44×10^{-36} e.s.u. for monomer and 14.32×10^{-36} e.s.u. for dimer) were found to be larger than those of standard organic compounds suggesting that 3ASA has a significant NLO character for optoelectronic applications. The NLO properties of dimer may differ from monomer due to dimerization. Further, the radiative lifetime, light harvesting efficiency and band gap energy were calculated, and proposed that 3ASA may be useful in photovoltaics and wide bandgap power devices.

Highlights

- DFT and TD-DFT theories were employed to calculate structural and molecular properties of 3ASA (monomer and dimer) in ground and excited states.
- HOMO-LUMO study shows monomer and dimer of 3ASA are good reactive.
- NBO analysis reflects that charge transfer interactions stabilized the 3ASA molecule.
- Electronic absorption/emission spectra in solvents calculated by IEF-PCM/TD-DFT method correlate with experimental results.
- Calculated NLO parameters suggested that 3ASA is a potential candidate for NLO material.

Keywords 3ASA · DFT/TD-DFT · Electronic spectra · NBO parameters · NLO properties · Dimer

Introduction

Salicylic acid (SA) is frequently used in the synthesis of organic compounds as well as in food preservation and plant hormones. It is extensively utilized in many skin care products to control hard skin, acne, corns, psoriasis, warts, and keratosis pilaris [1, 2]. It can even be used to treat headaches, arthritis pain, inflammation, and heart attack. It is an intermediate in the fabrications of agrochemicals, color products, and various dyes. Salicylic acid and its derivatives have drawn the interest of researchers due to their biological significance [1–5]. Moreover, 3-aminosalicylic acid is a

✉ Sanjay Pant
sanjayphotophys@gmail.com

✉ Mohan Singh Mehata
msmehata@gmail.com

[1] Photophysics Laboratory, Department of Physics, Centre of Advanced Study, DSB Campus, Kumaun University, Nainital 263002, India

[2] Laser Spectroscopy Laboratory, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

Published online: 02 July 2024

building block that can be utilized to synthesize antimycin A_3 [6].

Weller's crucial research on SA and its derivatives revealed that the photo-induced ESIPT (excited-state intramolecular proton transfer) reaction occurs between the nearby carboxylic group (COOH) and phenol group (OH) [7–9]. Based on computational and experimental examinations, it has been confirmed that ESIPT exists in salicylic acid derivatives in various solvents and polymers [7–15]. Jang et al. [16] have theoretically investigated the ESIPT mechanism for salicylic acid, 3-aminosalicylic acid (3ASA), 5-methoxy salicylic acid, and 5-aminosalicylic acid (5ASA) by potential energy curves calculated using the CIS (configuration interaction singles) and TD-DFT (time-dependent density functional theory) methods. Paul et al. [13–15] have studied the photophysics of various chloro-substituted salicylic acids (5-chlorosalicylic acid, 4-chlorosalicylic acid and 3,5,6-trichlorosalicylic acid) and explained the ESIPT process in various solvents. Joshi et al. [17] have studied the spectral and photophysical properties of substituted salicylic anions. Similar to the unsubstituted salicylic anion, the substituted anions show a large Stokes-shifted emission caused by ESIPT. Suyal et al. [18] have reported the photophysics of 4-aminosalicylic acid (4ASA) in polymer matrices of different polarity. These polymers displayed dual emission (shorter wavelength at 330 nm and longer wavelength at 410 nm in polyvinyl alcohol), with the bands at longer wavelengths being ascribed to ESIPT. Furthermore, Suresh et al. [1] have discussed the vibrational assignment, other quantum chemical properties and thermodynamic parameters of salicylic acid using Hartree Fock (HF) and DFT theories. Akkaya et al. [19] have reported the theoretical and experimental investigation of 4ASA and explored the modifications in the structure of dimers and various monomers. They also suggested that hydrogen bonding between carboxylic groups is accountable for dimerization. Singh et al. [20] have studied the experimental and theoretically calculated FT-IR/IR spectra of 3ASA, 4ASA, and 5ASA, which demonstrate different characteristics arising from low-frequency vibrational modes due to intramolecular motion. Moradi et al. [21] described the solubilization profile of 3ASA in the ethanol–water mixtures and it was observed that 3-ASA is preferentially solvated by water molecules in the intervals from neat water to the mixture x_1 (mole fraction of water) = 0.24 and from x_1 = 0.35 to neat ethanol. Nogueira et al. [22] have studied the intramolecular hydrogen bonding of 3-aminosalicylic acid on silver colloids using the surface-enhanced Raman (SER) spectra.

In recent years, computational calculations have been widely utilized to collect information regarding the molecular/electronic structures, orbital interactions, and other useful chemical properties of a molecule [23–39]. In quantum chemical computations, DFT and TD-DFT theories are very fruitful to predict the structural modifications, vibrational frequencies, and electronic spectra of a molecule [1, 3, 30, 34, 37]. DFT provides a good correlation between theoretical and experimental findings and facilitates prediction of molecular/electronic properties accurately [23–26, 35–38, 40]. These theories are also useful to obtain the kinetic stability, chemical reactivity, natural bond orbital (NBO) parameters and various thermodynamic parameters of a molecule [41–43]. Moreover, the NBO study of organic molecules is a comprehensive area for researchers to describe the numerous molecular properties such as bond type, bond order, occupancies and electron densities of donor–acceptor interactions, atomic charges, charge transfer or conjugative interactions as well as intramolecular and intermolecular interactions in a molecular system [1, 42, 44–47].

Additionally, it has been recognized that DFT is a straightforward and efficient approach to predict the nonlinear optical (NLO) properties of organic molecules such as polarizability, hyperpolarizability, and dipole moment [27, 28, 38, 39, 48–50]. The widespread use of NLO materials in photonics, telecommunications, frequency mixing, optical computing, electro-optical devices, molecular switches, dynamic holography, lasers, data storage, fast data transfer, medical diagnosis and other fields has made them an appealing and intriguing research topic. Because of their increased efficiency, greater molecule polarizability, and quick response time, organic NLO compounds offer several benefits over others [23, 51–53]. Moreover, the small HOMO (high occupied molecular orbital) – LUMO (lowest unoccupied molecular orbital) energy gap is essential for better NLO response. The HOMO–LUMO energy gap decreased by enhancing the conjugation. Structural modification is an effective approach to reduce the energy gap between the HOMO and LUMO [29, 30, 33, 34, 37–39].

This study focuses on the computational analysis of 3ASA, both as a monomer and a dimer, and examines its electronic and photophysical properties. Calculations include geometrical parameters, electronic absorption and emission spectra, HOMO–LUMO energy, energy gap, molecular electrostatic potential (MEP) map, and reactivity parameters. Additionally, NLO, NBO, and various thermodynamic parameters have been computed. Moreover, the spectral properties of 3ASA have been measured in solvents of varying polarity in order to compare the results with the theoretical calculations. Thus, the study could be valuable for applications in nonlinear optical materials and photovoltaics.

Computational and Experimental Details

The molecular structures of 3ASA were constructed using Gauss view 5.0 [54] and all the calculations were performed using the Gaussian 09 software [55]. The two conformers

(rotamer P and rotamer R) of 3ASA were constructed and optimized in the ground state in gas phase, by DFT/B3LYP/6-311G(d,p) method [24]. The B3LYP is a standard method, first developed to study vibrational absorption and circular dichroism, widely used and has a good compromise between computational cost, coverage, and accuracy of results [56]. Moreover, the higher basis set is more precise and more accurate for more complex molecular systems. 3-aminosalicylic acid is simple and thus, 6-311G(d,p) basis set is sufficient for the calculation. Frequency calculations were performed for the 3ASA (rotamer P and R) using the same method. This ensured that the molecular geometries were stable on the global minimum energy level because all frequencies are real (i.e. positive). The molecular structures of 3ASA (rotamer P and R) have also been optimized in the excited state. Further calculations such as HOMO–LUMO energy, MEP map, electronic transitions, NBO properties, NLO parameters, thermodynamic properties, and dimeric characteristics of 3ASA have also been performed for rotamer P. Moreover, TD-DFT theory [57] was employed to compute the electronic spectra (absorption/emission) of 3ASA in the gas phase. The effect of solvents on 3ASA has been studied using TD-DFT including the IEF-PCM (integral equation formalism of the polarizable continuum model) [58]. IEF-PCM is the current version of PCM useful in common quantum chemistry packages. One of the major challenges to PCM methods is that they are all derived assuming the solute charge density is entirely encapsulated in the cavity. This is often not the case, as the electron distributions frequently extend beyond the cavity. It has been demonstrated that IEF-PCM can effectively overcome this effect [59].

The NBO properties of 3ASA in the gas phase were calculated using NBO 3.1 program [24], which is incorporated in the Gaussian 09 software. The NLO parameters of 3ASA in the gas phase have been calculated by TD-DFT/B3LYP/6-311G(d,p) method. Furthermore, the thermodynamic properties of 3ASA at various temperatures were computed using the above mentioned theory in the gas phase. To study the dimeric nature of 3ASA, a molecular structure of a dimer of 3ASA was generated and optimized by DFT/B3LYP/6-311G(d,p) method in the ground and excited states and its properties viz. geometrical parameters, MEP, HOMO–LUMO energy and NBO parameters were obtained.

3ASA (purity: 97%) were purchased from Sigma-Aldrich and used as received. All the solvents used are of spectroscopic grade and de-ionized water/double distilled water was utilized whenever required. The absorption measurements were monitored using a UV/VIS/NIR spectrophotometer (Lambda-750, Perkin Elmer). The emission spectra were recorded with a Spectrofluorometer (Fluorolog-3 Horiba Jobin Yvon). A 450W xenon lamp was used as an excitation source. Transparent quartz cuvettes (path length: 1 cm) were used to measure the absorption and emission spectra.

Results and Discussion

Molecular Geometry Analysis

The molecular structures of 3ASA (rotamer P and R) were optimized using DFT and TD-DFT theories at B3LYP/6-311G(d,p) level in the ground and excited states, respectively. The optimized structures of 3ASA (rotamer P and R) are illustrated in Fig. 1 (a,b) including atomic numbering. The calculated energy of rotamer P is -551.57 a.u. in ground state and -551.45 a.u. in the excited state. For rotamer R, energy is -551.56 a.u. in the ground state and -551.44 a.u. in the excited state. The low energy values signify the stability of both the conformers. The comparative optimized geometrical parameters of 3ASA (for rotamer P & R) employing DFT/B3LYP/6-311G(d,p) are summarized in Tables 1 and 2. The optimized bond lengths and bond angles for 3ASA for both rotamers P and R are well matched with each other.

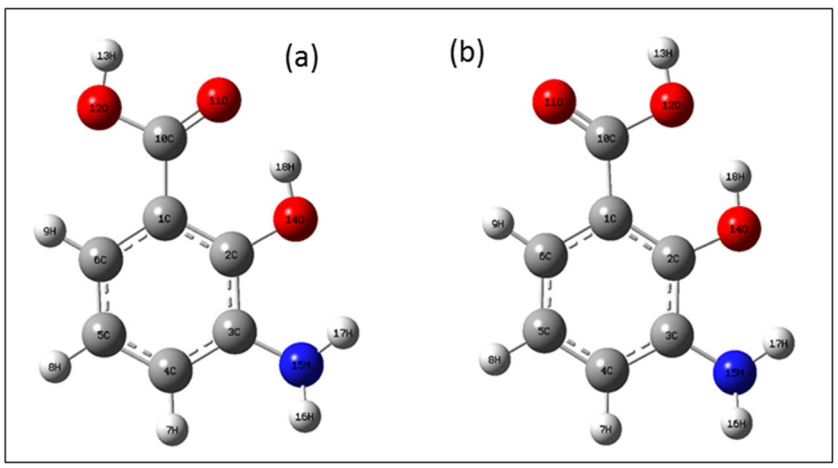


Fig. 1 Optimized structures of 3ASA (monomers) in the gas phase in the ground state (**a**) rotamer R, and (**b**) rotamer P with DFT/B3LYP/6-311G(d,p)

Table 1 Optimized bond lengths of 3ASA in the gas phase in the ground state (S_0) and excited state (S_1)

Bond length (Å)	3ASA DFT/B3LYP/6-311G(d,p)			SA* DFT/B3LYP/ 6311 + + G(d,p)	SA* (Exp)	3ASA TD-DFT/ B3LYP/ 6-311G(d,p)	
	R (S_0)	P (S_0)	Dimer (S_0)	(S_0)		R (S_1)	P (S_1)
C1-C2	1.41	1.41	1.41	1.41	1.40	1.42	1.42
C1-C6	1.41	1.41	1.41	1.40	1.40	1.40	1.40
C1-C10	1.48	1.46	1.47	1.50	1.48	1.45	1.45
C2-C3	1.42	1.42	1.42	1.40	1.39	1.42	1.42
C2-O14	1.36	1.35	1.35	1.36	1.36	1.35	1.33
C3-C4	1.39	1.39	1.39	1.39	1.38	1.42	1.43
C3-N15	1.39	1.39	1.39	-	-	1.36	1.36
C4-C5	1.40	1.40	1.40	1.39	1.40	1.38	1.37
C4-H7	1.09	1.09	1.09	1.08	1.01	1.08	1.08
C5-C6	1.38	1.38	1.38	1.39	1.39	1.41	1.42
C5-H8	1.08	1.08	1.08	1.08	1.15	1.08	1.08
C6-H9	1.08	1.08	1.08	1.09	1.08	1.08	1.08
C10-O11	1.20	1.23	1.25	1.20	1.23	1.24	1.28
C10-O12	1.38	1.35	1.32	1.37	1.31	1.43	1.37
O12-H13	0.97	0.97	1.00	0.96	1.01	0.96	1.50
O14-H18	0.97	0.98	0.98	0.96	0.82	1.69	0.97
N15-H16	1.01	1.01	1.01	-	-	0.99	1.04
N15-H17	1.01	1.01	1.01	-	-	1.01	1.01
O11·····H31	-	-	1.67				
H13·····O29	-	-	1.67				

* The values of bond length for salicylic acid (SA) were taken from the literature [1]

The computed bond lengths and bond angles of 3ASA (rotamer P and R) were compared with the reported theoretical (by DFT/B3LYP/6–311 + + G(d,p)) results of SA [1]. It is noticed that the calculated bond lengths and bond angles of 3ASA are in near correlation with the theoretical results of SA [1]. The optimized values (bond lengths and bond angles) related to the carboxylic group (COOH) slightly deviate from the reported value (Tables 1 and 2), which may be possible due to the inherent basis set deficiency at the B3LYP level. These optimized geometrical parameters are also compared to the reported experimental data of SA [1]. The computed bond lengths C4-C5 (1.40 Å), C6-H9 (1.08 Å) and C10-O11 (1.23 Å) of 3ASA exactly matched with experimental values of the parent molecule, i.e. salicylic acid (SA). Moreover, calculated bond angles C3-C4-C5 and C1-C10-O12 obtained the minimum deviation of 0.1° and 0.2°, respectively, when compared with the experimental ones (Tables 1 and 2). The calculated parameters of 3ASA nearly match the experimentally reported values. The maximum deviation in the value of bond length (O14-H18) ~ 0.16 Å and bond angle (C2-O14-H18) ~ 4.1° was observed in rotamer P. These deviations may probably be associated with the inter- or intra-molecular interactions in the solid phase of SA [1].

The optimized parameters of rotamer P and R of 3ASA in excited state are also given in Tables 1 and 2. Because of the translocation of electronic charge density in the excited state, the structural properties may be changed [24]. For rotamer P, the bond length O12-H13 increases by 0.53 Å, and C10-O11 increases by 0.05 Å whereas the bond angles C3-N15-H16 and H16-N15-H17 increase by 6°and 6.1°, respectively.

Additionally, it is probable that 3ASA forms intermolecular hydrogen-bonded complex in the solid phase and stabilizes the molecular system. To study the intermolecular hydrogen bonded complex formation, a dimeric structure of 3ASA (constructed from rotamer P) was optimized in the ground state (gas phase) using DFT/B3LYP/6-311G(d,p) method. The optimized structure of the dimer is shown in Fig. 2 and some significant parameters are given in Table 1 and Table 2. The dimeric structure is more stable than the monomeric structure because the calculated energy (energy = -1103.16 a.u.) is approximately double in comparison to the monomer (rotamer P: energy = -551.57 a.u.). The stability of dimer may arise due to the formation of intermolecular hydrogen bonds between the atoms of the carboxylic group (between atoms O11 & H31, and O29 & H13) (Fig. 2) [19, 44]. The calculated bond lengths and bond angle values of the dimer are closer to that of rotamer P. The bond length

Table 2 Optimized bond angles of 3ASA in a gas phase in the ground and excited states

Bond Angle (°)	3ASA DFT/B3LYP/6-311G(d,p)			SA* DFT/B3LYP/6311 + + G(d,p)	SA* (Exp)	3ASA TD-DFT/B3LYP/6-311G(d,p)	
	R (S_0)	P (S_0)	Dimer (S_0)	(S_0)		R (S_1)	P (S_1)
C2-C1-C6	119.7	120.1	119.9	118.9	119.8	115.3	116.6
C2-C1-C10	124.0	118.0	119.3	121.0	118.9	125.2	117.6
C6-C1-C10	116.3	121.9	120.8	120.0	121.4	119.4	125.8
C1-C2-C3	120.0	120.2	120.2	119.7	119.7	120.9	120.7
C1-C2-O14	125.4	123.7	124.0	118.2	122.4	123.4	121.0
C3-C2-O14	114.5	116.1	115.8	122.9	117.9	115.7	118.4
C2-C3-C4	118.8	118.4	118.4	120.5	119.6	121.8	121.2
C2-C3-N15	118.5	118.4	118.4	-	-	117.7	117.5
C4-C3-N15	122.7	123.2	123.2	-	-	120.5	121.3
C3-C4-C5	121.3	121.5	121.4	120.4	121.6	117.1	117.8
C3-C4-H7	118.8	118.8	118.8	119.4	-	120.2	120.0
C5-C4-H7	120.0	119.8	119.8	120.2	-	122.7	122.2
C4-C5-C6	120.1	120.4	120.3	119.2	118.8	121.0	120.9
C4-C5-H8	119.6	119.4	119.5	120.6	-	120.3	120.3
C6-C5-H8	120.3	120.2	120.2	120.2	-	118.7	118.8
C1-C6-C5	120.2	119.6	119.7	121.4	120.5	123.7	122.8
C1-C6-H9	118.0	119.1	118.9	119.1	-	116.5	118.2
C5-C6-H9	121.8	121.3	121.4	119.5	-	119.7	119.0
C1-C10-O11	126.5	124.7	122.3	125.0	122.2	127.6	122.8
C1-C10-O12	113.5	114.8	116.0	114.8	115.0	112.4	116.5
O11-C10-O12	120.0	120.5	121.8	120.2	122.9	119.9	120.6
C10-O12-H13	106.7	106.3	110.4	109.9	-	105.4	105.0
C2-O14-H18	109.1	107.3	107.3	-	111.4	107.5	104.0
C3-N15-H16	115.5	115.7	115.7	-	-	121.3	121.7
C3-N15-H17	114.3	114.3	114.3	-	-	119.2	118.6
H16-N15-H17	113.5	113.6	113.6	-	-	119.5	119.7

* The values of bond angle for salicylic acid (SA) were taken from the literature [1]

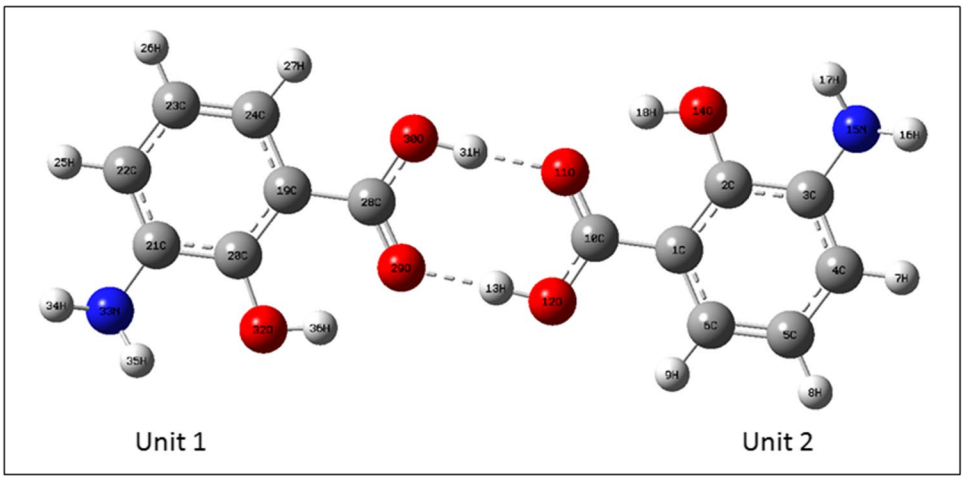


Fig. 2 Optimized molecular structure of dimeric form of 3ASA in gas phase with DFT/B3LYP/6-311G(d,p)

C10-O12 decreases by 0.3 Å and the bond length O12-H13 increases by 0.3 Å upon dimerization. These changes perhaps occurred due to the delocalization of the electronic charge densities on the electro-negative atom O12. The bond angle C10-O12-H13 increases by 4.1˚and the bond angle C1-C10-O11 reduces by 2.4˚ upon dimerization when compared with rotamer P. Thus, it can be inferred that intermolecular hydrogen bonding interactions are responsible for modifying the molecular structure of dimer and corresponding parameters [44]. Moreover, the cartesian coordinates of the geometric optimization for both the monomer and dimer forms of 3ASA were calculated. The cartesian coordinates for monomers P and R in the gas phase, in the ground and excited states, are detailed in Tables S1 and S2 of the supplementary information (SI). Furthermore, the Cartesian coordinates pertaining to the geometric optimization of the dimer in the gas phase, specifically in the ground state, have been compiled in Table S3 of the SI.

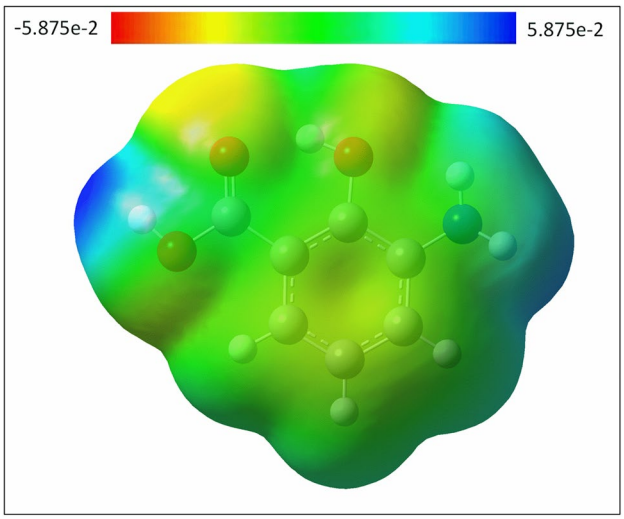


Fig. 3 MEP map of 3ASA (rotamer P) obtained using DFT/B3LYP/6-311G(d,p) in the gas phase

The association between the carboxylic groups of two monomers (i.e., rotamer P) results in stronger intermolecular hydrogen bonding (O11⋯⋯H31 (1.67 Å) and O29-H13 (1.67 Å)) in the dimer. These calculated bond lengths were estimated within the range of hydrogen bonding interactions (< 3.0 Å). Thus, the hydrogen bonding stabilizes the molecular system by pushing the two heteroatoms closer to one another [3].

Molecular Electrostatic Potential (MEP) Map

The MEP map of 3ASA (rotamer P) was generated using the DFT level and is shown in Fig. 3 with a color range of -5.875e-2 a.u. to 5.875e-2 a.u. The electrostatic potential is characterized by various colors, potential values in increasing order are arranged as red < yellow < green < blue. The portions of maximum and minimum potential are considered electron-rich and electron-deficient, respectively [3]. In Fig. 3, the maximum yellow color region (maximum negative electrostatic potential) mainly around the oxygen atom (O11) is electron-rich and it is a preferred region for an electrophilic attack [24, 44]. In the blue color region; positive electrostatic potential is mainly over the hydrogen atoms H13, H16, and H17, which are electron-deficient and provide the area for the attack of nucleophilic species. The region around H13 (Fig. 3) is the bluest and it is a predominant area of a nucleophilic attack [42, 60]. The MEP map is very convenient in the exploration of biological recognition mechanisms and intermolecular hydrogen bonding interactions [23, 61].

Figure 4 represents the MEP map for dimer within the color range from -4.463e-2 a.u. to 4.463e-2 a.u. The atoms

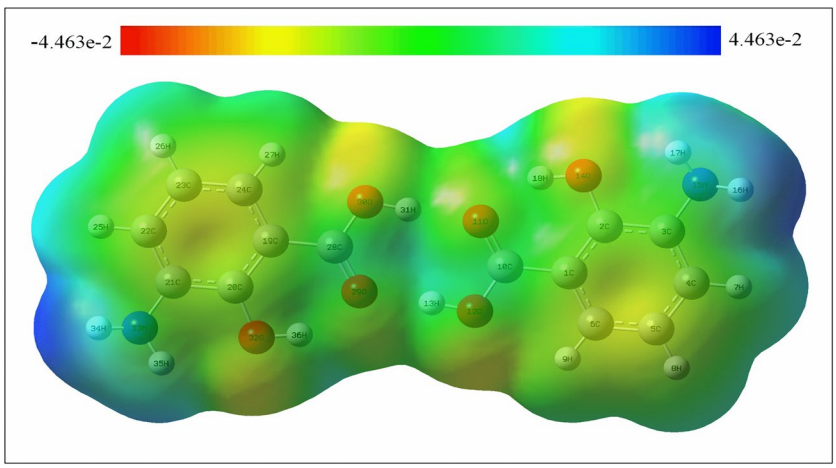


Fig. 4 MEP map of dimer (3ASA) obtained using DFT/B3LYP/6-311G(d,p) in gas phase

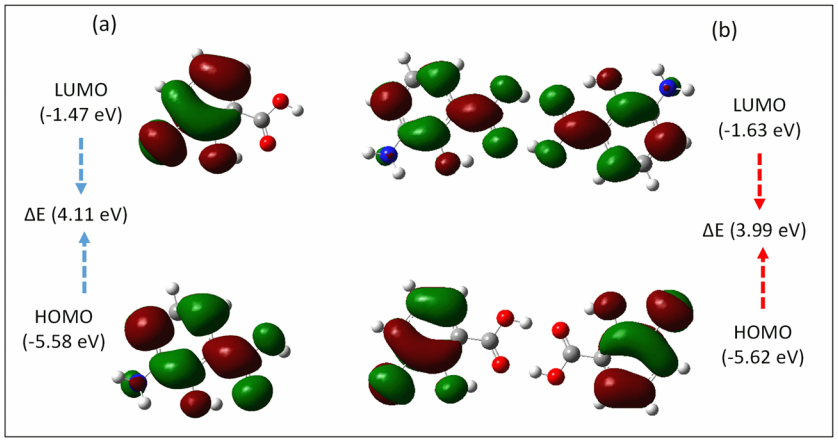


Fig. 5 HOMO–LUMO of 3ASA (**a**) monomer and (**b**) dimer obtained using DFT/B3LYP/6-311G(d,p)

Table 3 Computed HOMO–LUMO energy, energy gap, and reactivity parameters of 3ASA (monomer and dimer) using DFT/B3LYP/6-311G(d,p)

Parameters	Monomer	dimer
E_{HOMO}	-5.58 (eV)	-5.62 (eV)
E_{LUMO}	-1.47 (eV)	-1.63 (eV)
$\Delta E = (E_{HOMO} - E_{LUMO})$	4.11 (eV)	3.99 (eV)
$I = -E_{HOMO}$	5.58 (eV)	5.62 (eV)
$A = -E_{LUMO}$	1.47 (eV)	1.63 (eV)
$\chi = (I + A)/2,$	3.53 (eV)	3.63 (eV)
$\mu = -(I + A)/2$	-3.53 (eV)	-3.63 (eV)
$\eta = (I - A)/2$	2.06 (eV)	2.00 (eV)
$S = 1/\eta$	0.49 (eV^{-1})	0.50 (eV^{-1})
$\omega = \mu^2/2\eta$	3.07 (eV)	3.29 (eV)
$\omega^+ = (I + 3A)^2/16(I - A)$	1.52 (eV)	1.73 (eV)
$\omega^- = (3I + A)^2/16(I - A)$	5.04 (eV)	5.36 (eV)

The reactivity parameter relations given in Table 3 used from the references [23, 24]

O11 and H13 (of monomer unit 1) strongly interact with the H31 and O29 atoms (of monomer unit 2), respectively, through intermolecular hydrogen bonding. Thus, the dimer is more stable than the monomer due to strong intermolecular hydrogen bonding [19, 44]. Due to dimer formation, now, the region around the atoms O11, H13, O29 and H311 is green, which signifies the zero electrostatic potential regions. The yellow color is located around the atoms O12, O14, O32 and O38, and these are the electrophilic regions, and the hydrogen atoms mainly associated with the amino groups are blue and are nucleophilic regions. Through MEP map analysis, it can be observed that the electrophilic and nucleophilic regions are different in monomer and dimer, and their molecular and electronic properties may also vary.

Frontier Molecular Orbitals (FMOs) and Chemical Reactivity Parameter Analysis

The FMOs provide information about chemical reactivity, kinetic stability, electronic spectra, electrical, and optical properties of the molecules [24, 60]. Figure 5 (a,b) illustrates the highest occupied and lowest unoccupied molecular orbitals (HOMO and LUMO) for 3ASA (rotamer P) monomer and dimer (in the gas phase obtained by DFT/B3LYP/6-311G(d,p)). The HOMO LUMO energy and its ΔE values are summarized in Table 3.

HOMO signifies electron donor and LUMO acts as an electron acceptor [60, 61]. In monomer and dimer, HOMO is positioned over amino and hydroxyl groups as well as the benzene ring, whereas LUMO is located over the whole 3ASA moiety. The value of energy gap ΔE is 4.11 eV (for monomer) is small, and is responsible for the charge transfer interactions within the molecular system [23, 61]. The value of ΔE is 3.99 eV for the dimer, which signifies that it is more reactive and stable than the monomer [62].

Additionally, various quantum chemical reactivity descriptors such as ionization potential (A), electron affinity (I), chemical potential (μ), electronegativity (χ), softness (S), hardness (η), and electrophilicity indices (ω), electron accepting/donating power (ω^+, ω^-) [23, 24, 61] were estimated using HOMO–LUMO energy for monomer and dimer in the gas phase and are listed in Table 3. The I and A values signify the electron donating and accepting power of a probe. The large value of I (5.58 eV) and ω^- (5.04 eV) indicate that 3ASA (monomer) has electron-donating nature. The positive value of A (1.47 eV) signifies the charge transfer character within the 3ASA molecule. The negative μ (-3.53 eV) value indicated that 3ASA is quite stable. The η and S are associated with the stability and reactivity of the molecule. The large η (2.06 eV) value shows that 3ASA is

Table 4 Electronic absorption and emission transitions and other corresponding parameters of 3ASA in gas/solvent phases obtained with TD-DFT/B3LYP/6-311G(d,p)

Absorption results							Emission results	
Solvent	Electronic transitions	E_{ex} (eV)	λ_{abs} (nm)	*f*	assignment	CI value	λ_{em} (nm)	*f*
GAS	$S_0 \rightarrow S_1$	3.61	343	0.06	H→L	0.70	429	0.05
TOL	$S_0 \rightarrow S_1$	3.58	347	0.08	H→L	0.70	435	0.07
DEE	$S_0 \rightarrow S_1$	3.58	346	0.07	H→L	0.70	438	0.08
THF	$S_0 \rightarrow S_1$	3.58	346	0.08	H→L	0.70	441	0.09
ACN	$S_0 \rightarrow S_1$	3.59	345	0.07	H→L	0.70	444	0.10
DMSO	$S_0 \rightarrow S_1$	3.58	346	0.08	H→L	0.70	448	0.10
MeOH	$S_0 \rightarrow S_1$	3.59	345	0.07	H→L	0.70	444	0.10
Water	$S_0 \rightarrow S_1$	3.59	345	0.07	H→L	0.70	445	0.10

[#] HOMO→H and LUMO→L

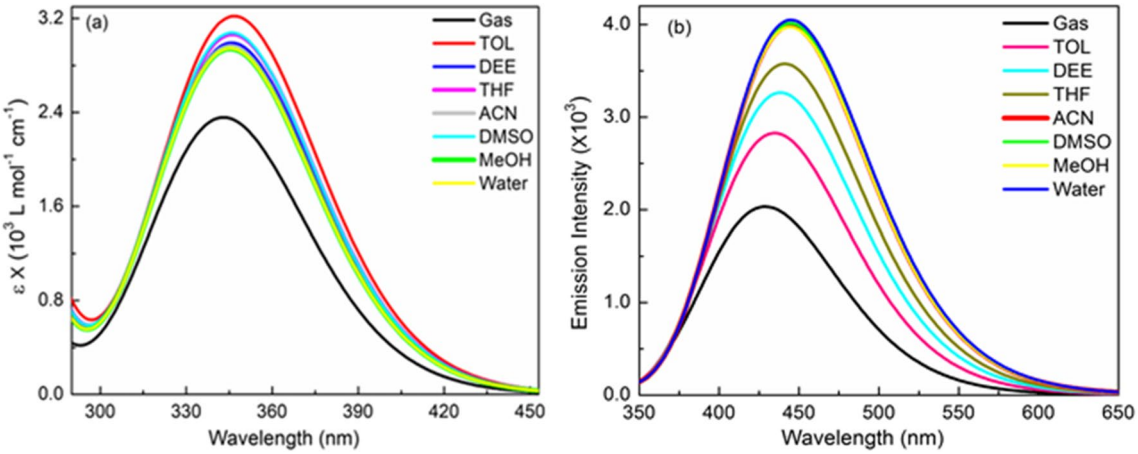


Fig. 6 Simulated (**a**) absorption and (**b**) emission spectra of 3ASA in gas/solvent phases using TD-DFT/B3LYP/6-311G(d,p) method

stable, and a moderate value of S (0.49 eV^{-1}) indicates that 3ASA is moderately reactive. The molecular structure, bonding interactions, reactivity, and stability of the molecule are described with ω value. The higher ω (3.07 eV) value shows that 3ASA is a good electrophile as $\omega > 1.5$ [24]. Furthermore, the greater μ, S and ω values propose that the dimeric structure of 3ASA is more reactive and stable as compared to the monomeric form [62].

Electronic Absorption and Emission Spectra

Electronic absorption and emission (fluorescence) spectra of a monomer (rotamer P of 3ASA) in the gas phase and different solvents were calculated using the TD-DFT/B3LYP method with a 6-311G(d,p) basis set. To investigate the influence of solvents on the 3ASA monomer, TD-DFT was combined with an IEF-PCM solvation model. A series of solvents of different polarities were selected, including non-polar, polar aprotic and polar protic solvents. This selection enables the investigation of the solvent effect on the photophysical properties of 3ASA, providing insights into

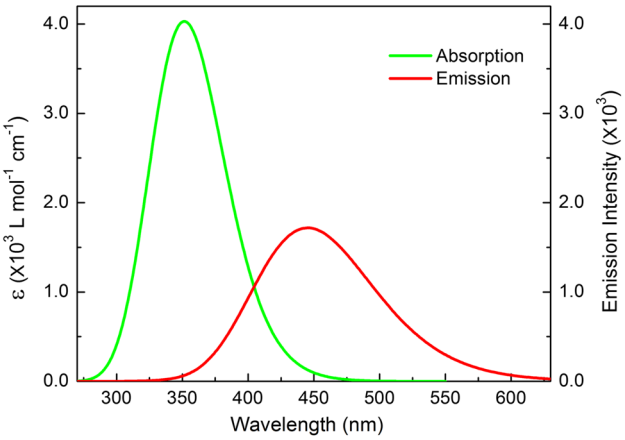


Fig. 7 Simulated absorption and emission spectra of dimer in the gas phase using TD-DFT/B3LYP/6-311G(d,p) method

how different solvent environments influence its behavior. The various parameters in the ground state such as excitation energy (E_{ex}), absorption wavelength λ_{abs} (nm), orbital assignment, CI value, and oscillator strength *(f)* of 3ASA

Table 5 Experimentally observed spectral parameters of 3ASA in solvents

Solvents	λ_{abs} (nm)	λ_{em} (nm)	ν_{abs} (cm^{-1})	ν_{em} (cm^{-1})	$\Delta\nu$ (cm^{-1})	f	τ_0 (ns)	LHE
DEE	342	411	29,240	24,331	4909	0.06	29.2	0.13
THF	344	418	29,070	23,923	5147	0.05	35.5	0.11
ACN	340	427	29,412	23,419	5993	0.08	21.7	0.17
MeOH	322	400	31,056	25,000	6056	0.06	25.9	0.13
Water	302	407	33,113	24,570	8543	0.05	27.4	0.11

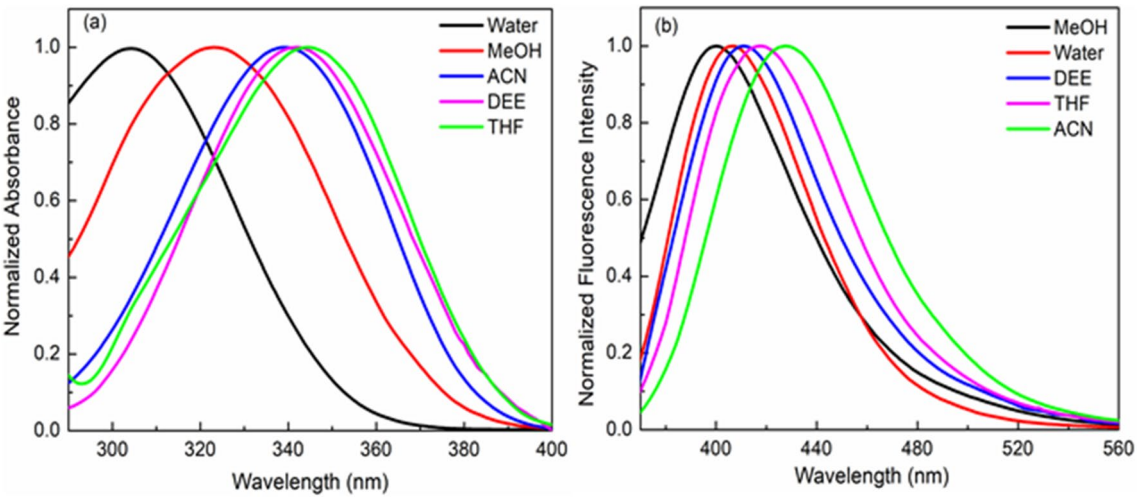


Fig. 8 Experimentally observed normalized (**a**) absorption and (**b**) emission spectra (excited at corresponding absorption wavelength) of 3ASA in different solvents

in the gas/solvent phases are listed in Table 4. The strong electronic absorption band of 3ASA was primarily obtained from the inclusion of HOMO → LUMO with a maximum contribution of 98% (CI value: 0.70). Thus, the main absorption band appeared in the range of 343–347 nm in the gas phase and different solvents. The calculated electronic absorption and emission spectra of 3ASA (monomer P) are shown in Fig. 6 (a,b) in gas/solvent phases. The absorption band of 3ASA was predicted around 343 nm in the gas phase, 347 nm in toluene (TOL), 346 nm in diethyl ether (DEE), 346 nm in tetrahydrofuran (THF), 345 nm in acetonitrile (ACN), 345 nm in methanol (MeOH), 346 nm in DMSO, and 345 nm in water with 98% contribution of HOMO → LUMO.

The emission spectra of 3ASA (Fig. 6 (b)) appeared around 429 nm in gas, 435 nm in TOL, 438 nm in DEE, 441 nm in THF, 444 nm in ACN, 444 nm in MeOH, 448 nm in DMSO and 445 nm in water. Thus, it can be noticed that absorption and emission spectra are red shifted in solvents as compared to the gas phase due to the solvent effect [50]. Moreover, the electronic spectra of the dimer were also calculated theoretically in the gas phase using the same method and shown in Fig. 7 (a,b). The computed absorption (λ_{abs} = 345 nm) and emission (λ_{em} = 445 nm) spectra of dimer are consistent with the theoretically observed spectra of the monomer (in water).

To authenticate the computed spectral properties, absorption/emission spectra of 3ASA (at 0.1 mM) were recorded experimentally in some solvents (Table 5). Figure 8 (a,b) shows the absorption spectra in the range of 290–400 nm and emission spectra of 3ASA in some solvents. It is important to state that the 3ASA is less soluble in non-polar solvents. Therefore, the absorption spectrum was not observed in TOL for the chosen concentration with the available instrument. The absorption spectra of 3ASA were appeared at 340 in ACN, 342 nm in DEE, 344 nm in THF, 322 nm in MeOH, and 302 nm in water. The emission maxima were observed at 411 nm in DEE, 418 nm in THF, 427 nm in ACN, 400 nm in MeOH, and 407 nm in water. Due to low solubility, the clear emission peak of 3ASA does not appear in TOL. Both spectra (absorption/emission) of 3ASA in water and MeOH are blue-shifted as compared to non-polar/polar protic solvents due to hydrogen bonding/anion formation (deprotonation of COOH) [23, 63, 64].

The absorption band maxima in DEE, THF, and ACN nearly correspond to the experimental ones. The absorption peak shows variations of 23 and 43 nm in MeOH and water, respectively, as compared to the theoretical values.

The difference in the theoretical and experimental values of emission maxima (in solvents) varies from 17 - 44 nm. The variations in experimental and theoretical results of absorption and emission spectra may be possible because experimental observations depend on various physical parameters such as the temperature of the surroundings, the concentration of solute and hydrogen bonding ability of solvents, etc., whereas in computational calculations solute–solvent interactions are not considered [65, 66].

The fluorophores that possess charge transfer characteristics are very useful in optoelectronic devices and are greatly influenced by the radiative (or excited state) lifetime. It is expected that fluorophores with considerably longer lifetimes will demonstrate effective electron injection and charge transfer [67]. The radiative lifetime of the fluorophore is calculated by Eq. 1.

$$\tau_0 = \frac{1.5}{f \times \nu_{abs}^2 (cm^{-1})} \tag{1}$$

where ν_{abs} is the absorption wavenumber. f is the oscillator strength [68] and can be calculated using Eq. 2,

$$f = 4.33 \times 10^{-9} \left[\int \epsilon_\nu d\nu\right]_{exp} \tag{2}$$

where ε_ν is the molar extinction coefficient at wavenumber (ν) measured from the absorption spectrum. $\left[\int \epsilon_\nu d\nu\right]_{exp}$ is the integrated area under the absorption spectrum.

The computed results (Table 5) show that 3ASA has large τ_0 values, resulting in greater photocurrent efficiency in photovoltaics, particularly dye-sensitized solar cells [67].

Light harvesting efficiency (LHE) predicts the ability of organic compounds to absorb photons and then inject photoexcited electrons into the conduction band of semiconductors [67]. LHE is estimated using the Eq. 3 and is given in Table 5.

$$LHE(\lambda) = 1 - 10^{-f} \tag{3}$$

The LHE values (experimentally estimated) of 3ASA have been found in the range of 0.11–0.17 in solvents which is close to the LHE value (0.21) of Red mulberry (H+) [69]. Moreover, the experimentally observed LHE values well matched with theoretically calculated LHE values (0.15–0.17) (Table 5). The result shows that 3ASA can act as a good candidate for light-harvesting materials which are significant for photoacclimation, photoprotection and photovoltaics [67, 69, 70].

Band gap energy (E_g) is an informative parameter to discuss the kinetic stability, chemical reactivity and optical polarizability [24, 71], and is calculated using Eq. 4

$$E_g = \frac{hc}{\lambda_{ae}} = \frac{1240}{\lambda_{ae}} \tag{4}$$

where λ_{ae} denotes the absorption edge wavelength (nm), h indicates Planck's constant and c is the speed of light.

The E_g value of 3ASA was calculated (experimentally) to be 3.18 eV in THF solvent. The calculated E_g value of 3ASA is close to the E_g values of wide bandgap power semiconductors such as gallium nitride (GaN: 3.39 eV) and 4H-silicon carbide (4H-SiC: 3.32 eV) and 6H-silicon carbide (6H-SiC: 2.9 eV) [72]. Wide bandgap power devices are integrated as replacements for silicon technology in multiple technological applications such as electric vehicle (EV) charging, motor drive applications, renewable energy systems, etc. [73].

Natural Bond Orbital (NBO) Analysis

The NBO calculations for 3ASA (rotamer P) were done by employing DFT/B3LYP/6-311G(d,p) method. The estimated hybridization and their corresponding characters of NBOs for 3ASA are tabulated in Table 6. The atoms O and N are considered to be highly electro-negative as the P_c (polarization coefficient square) value is large [23]. The other atoms connected with these O and N atoms have small electronegativity due to low P_c values. For example, the NBO σ(N15-H16) is established by the contribution of 68.47% of the electron density of the N15 and 31.53% of the H16 atom. This NBO is originated by the connection of $sp^{2.93}$ ($P_c = 0.83$ and p-character: 74.51%) of the N15 atom and $sp^{0.00}$ ($P_c = 0.56$ and p-character: 0.06%) of the H16 atom.

When the donor–acceptor interactions arise within the molecular framework, the occupancies and energies of donor orbitals reduce and charge transfer interactions occur [23, 44]. The charge transfer interactions between donor and acceptor NBO can be described by the stabilization energy ($E^{(2)}$) and corresponding relation given by Eq. 5.

$$E^{(2)} = \Delta E_{ij} = (q_i F^2(i,j)/(\epsilon_j - \epsilon_i)), \tag{5}$$

here, $E^{(2)}$ is the stabilization energy and is computed using the second-order perturbation theory. q_i is the orbital occupancy of the donor. $\in_i$ and $\in_j$ correspond to diagonal elements, and $F(i, j)$ signifies the off-diagonal NBO Fock matrix element.

A larger stabilization energy indicates stronger donor–acceptor interactions and a higher degree of conjugation throughout the entire molecular system [24]. Table 7 shows the stabilization energy for a few significant donor–acceptor interactions. The most significant NBOs are π(C1-C2)→π*(C10-O11), LP(2) O14→π*(C1-C2) and LP (2) O12→π*(2) (C10—O11) with higher $E^{(2)}$ values 29.54, 33.84 and 46.76 kcal/mole, respectively. The NBOs interactions probably arise because of the delocalization of π electrons from one NBO to another and hence charge transfer

Table 6 Bond type, occupancy, electron density, p character of significant natural atomic hybrid of the NBO of 3ASA (rotamer P)

Bond (A-B)	Occupancy	ED_A (%)	ED_B (%)	Hybrid	Atom	p (%)
σ(C1-C2)	1.97	50.85	49.15	$0.71sp^{1.95}+0.70sp^{1.58}$	C1	66.04
					C2	61.26
π(C1-C2)	1.66	56.92	43.08	$0.75sp^{1.00}+0.66sp^{1.00}$	C1	99.98
					C2	99.95
σ(C1-C10)	1.98	51.63	48.37	$0.72sp^{2.36}+0.70sp^{1.46}$	C1	70.20
					C10	59.31
σ(C2-O14)	1.99	33.17	66.83	$0.58sp^{2.99}+0.82sp^{1.92}$	C2	74.80
					O14	65.67
σ(C3-N15)	1.99	40.53	59.47	$0.64sp^{2.49}+0.77sp^{1.64}$	C3	71.32
					N15	62.15
π(C10-O11)	1.99	27.10	72.90	$0.52sp^{1.00}+0.85sp^{1.00}$	C10	99.43
					O11	99.89
σ(O12-H13)	1.99	74.56	25.44	$0.86sp^{3.82}+0.50sp^{0.00}$	O12	79.20
					H13	99.83
σ(N15-H16)	1.99	68.47	31.53	$0.83sp^{2.93}+0.56sp^{0.00}$	N15	74.51
					H16	0.06
LP(1) O11	1.97			$sp^{0.70}$	O11	41.29
LP(2) O11	1.84			$sp^{91.14}$	O11	98.86
LP(1) O12	1.98			$sp^{1.24}$	O12	55.35
LP(2) O12	1.82			$sp^{1.00}$	O12	99.94
LP(1) O14	1.97			$sp^{1.40}$	O14	58.30
LP(2) O14	1.84			$sp^{1.00}$	O14	99.94
LP(1) N15	1.84			$sp^{8.32}$	N15	89.23
π^*(C1-C2)	0.44	43.08	56.92	$0.66sp^{1.00}$–$0.75sp^{1.00}$	C1	99.98
					C1	99.95
π^*(C3-C4)	0.37	51.00	49.00	$0.71sp^{1.00}$ -$0.70sp^{99.99}$	C4	99.94
					C5	99.95
π^*(C10-O11)	0.31	72.90	27.10	$0.85sp^{1.00}$ -$0.52sp^{1.00}$	C10	99.43
					O11	99.89

Table 7 Stabilization energies for some significant donor–acceptor interactions of 3ASA

Donor NBO (i)	Acceptor NBO (j)	E^2 (kcal/mol)	E(j)-E(i) (a.u.)	F(i,j) (a.u.)
π(C1-C2)	π*(C3-C4)	15.83	0.30	0.062
π(C1-C2)	π*(C5-C6)	19.50	0.30	0.069
π(C1-C2)	π*(C10-O11)	29.54	0.25	0.077
π(C3-C4)	π*(C1-C2)	20.91	0.28	0.070
π(C3-C4)	π*(C5-C6)	18.98	0.30	0.067
π(C5-C6)	π*(C1-C2)	17.05	0.27	0.063
πC5-C6)	π*(C3-C4)	18.85	0.28	0.067
LP(2) O11	σ*(C10-O12)	29.13	0.65	0.125
LP(2) O11	σ*(O14-H18)	15.19	0.71	0.095
LP(2) O12	π*(C10-O11)	46.76	0.33	0.115
LP(2) O14	π*(C1-C2)	33.83	0.34	0.103
LP(1) N15	π*(C3-C4)	27.78	0.32	0.089

interactions occur within the 3ASA molecule, which stabilizes the molecular framework.

The NBO study was performed in different solvents with the same method and related parameters are collected in Table 8. The $E^{(2)}$ values of various NBO differ may be due to the translocation of charge in a distinct degree due to solvent–solute interactions. The greater $E^{(2)}$ values of 47.96, 47.66 and 47.39 kcal/mol were estimated for the NBO LP(2)

Table 8 Stabilization energies for some significant donor–acceptor interactions of 3ASA in different solvents

Donor NBO (i)	Acceptor NBO (j)	$E^{(2)}$ (kcal/mol)	E(j)-E(i) (a.u.)	F(i,j) (a.u.)
ACN				
π(C1-C2)	π*(C3-C4)	15.72	0.30	0.062
π(C1-C2)	π*(C5-C6)	19.27	0.31	0.069
π(C1-C2)	π*(C10-O11)	29.43	0.25	0.077
π(C3-C4)	π*(C1-C2)	21.23	0.27	0.070
π(C3-C4)	π*(C5-C6)	19.06	0.30	0.068
π(C5-C6)	π*(C1-C2)	17.33	0.27	0.067
πC5-C6)	π*(C3-C4)	18.82	0.28	0.067
LP(2) O11	σ*(C10-O12)	28.72	0.65	0.125
LP(2) O11	σ*(O14-H18)	15.44	0.71	0.096
LP(2) O12	π*(C10-O11)	47.96	0.33	0.115
LP(2) O14	π*(C1-C2)	33.24	0.34	0.103
LP(1) N15	π*(C3-C4)	27.88	0.32	0.089
MeOH				
π(C1-C2)	π*(C3-C4)	15.78	0.30	0.062
π(C1-C2)	π*(C5-C6)	19.54	0.30	0.069
π(C1-C2)	π*(C10-O11)	29.26	0.25	0.077
π(C3-C4)	π*(C1-C2)	21.07	0.27	0.070
π(C3-C4)	π*(C5-C6)	19.34	0.30	0.068
π(C5-C6)	π*(C1-C2)	17.05	0.27	0.063
πC5-C6)	π*(C3-C4)	18.52	0.29	0.066
LP(2) O11	σ*(C10-O12)	28.82	0.65	0.125
LP(2) O11	σ*(O14-H18)	15.33	0.71	0.096
LP(2) O12	π*(C10-O11)	47.66	0.33	0.115
LP(2) O14	π*(C1-C2)	33.69	0.34	0.103
LP(1) N15	π*(C3-C4)	28.28	0.32	0.090
Water				
π(C1-C2)	π*(C3-C4)	15.83	0.30	0.062
π(C1-C2)	π*(C5-C6)	19.81	0.30	0.069
π(C1-C2)	π*(C10-O11)	29.09	0.25	0.077
π(C3-C4)	π*(C1-C2)	20.91	0.28	0.070
π(C3-C4)	π*(C5-C6)	19.63	0.29	0.068
π(C5-C6)	π*(C1-C2)	16.79	0.27	0.063
πC5-C6)	π*(C3-C4)	18.21	0.29	0.066
LP(2) O11	σ*(C10-O12)	28.91	0.65	0.125
LP(2) O11	σ*(O14-H18)	15.23	0.71	0.095
LP(2) O12	π*(C10-O11)	47.39	0.33	0.115
LP(2) O14	π*(C1-C2)	34.13	0.34	0.103
LP(1) N15	π*(C3-C4)	28.77	0.32	0.090

Table 9 Stabilization energies for some significant donor–acceptor interactions of a dimer

Donor NBO (i)	Acceptor NBO (j)	E^2 (kcal/mol)	E(j)-E(i) (a.u.)	F(i,j) (a.u.)
Unit 1 → Unit 2				
LP(1) O11	σ*(O30-H31)	10.79	1.11	0.098
LP(2) O11	σ*(O30-H31)	15.86	0.70	0.096
Unit 2 → Unit 1				
LP(1) O29	σ*(O12-H13)	10.79	1.11	0.098
LP(2) O29	σ*(O12-H13)	15.86	0.70	0.096

O12 → π*(C10 – O11) in ACN, MeOH and water, respectively. The larger $E^{(2)}$ values of NBOs in solvents than in the gas phase reflect that the charge transfer interactions are stronger in solvents than in the gas phase due to solvent–solute interactions. Moreover, in dimer, the $E^{(2)}$ = 10.79 kcal/mole was found for the NBO LP(1) O11 → σ*(O30-H31) and LP(1)O29 → σ*(O12-H13) whereas the $E^{(2)}$ = 15.86 kcal/mole has been obtained for NBO LP(2) O11 → σ*(O30-H31) and LP(2) O29 → σ*(O12-H13) (Table 9). These larger values of $E^{(2)}$ suggest that intermolecular hydrogen bonding between carboxylic groups is responsible for dimerization [3, 44].

Non-Linear Optical (NLO) Analysis

The NLO properties: polarizability (α), first-order hyperpolarizability (β) and second-order hyperpolarizability (γ) of 3ASA were calculated using DFT/B3LYP/6-311G(d,p) method. The polarizability (α), first-order hyperpolarizability (β) and second-order hyperpolarizability (γ) values of 3ASA are calculated using the following Eqs. (6–11) [23, 24, 74].

$$\alpha_{total} = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3 \tag{6}$$

$$\beta_{total} = \left(\beta_x^2 + \beta_y^2 + \beta_z^2\right)^{1/2} \tag{7}$$

where,

$$\beta_x = \beta_{xxx} + \beta_{xyy} + \beta_{xzz} \tag{8}$$

$$\beta_y = \beta_{yyy} + \beta_{yzz} + \beta_{yxx} \tag{9}$$

$$\beta_z = \beta_{zzz} + \beta_{zxx} + \beta_{zyy} \tag{10}$$

$$\gamma_{total} = \frac{1}{5}\left[\left(\gamma_{xxxx} + \gamma_{yyyy} + \gamma_{zzzz}\right) + 2\left(\gamma_{xxyy} + \gamma_{xxzz} + \gamma_{yyzz}\right)\right] \tag{11}$$

where α is the polarizability tensor. The β and γ are the first-order and second-order tensors, respectively. Moreover, the first-order hyperpolarizability is represented as a 3 × 3 × 3 matrix and is a tensor of rank 3. According to Kleinman symmetry, the twenty-seven components of the 3 × 3 × 3 matrix can be diminished to ten components [24, 75]. The components of α (α_{xx}, α_{xy}, α_{yy}, α_{xz}, α_{zy}, and α_{zz}), β (β_{xxx}, β_{xxy}, β_{xyy}, β_{yyy}, β_{xxz}, β_{xyz}, β_{yyz}, β_{xzz}, β_{yzz} *and* β_{zzz}) and γ (γ_{xxxx}, γ_{xxyy}, γ_{yyyy}, γ_{xxzz}, γ_{yyzz}, γ_{zzzz}) have been obtained from Gaussian output.

The components of α and β for 3ASA (monomer and dimer) are summarized in Table 10 and the calculated α_{total} and β_{total} values are collected in Table 11 and compared with the reported data. The calculated values of α_{total} and β_{total} for 3ASA are 13.86×10^{-24} e.s.u. and 4.21×10^{-30} e.s.u., respectively. The α_{total} value of 3ASA is slightly small when compared to acetylsalicylic acid [51], 3,5-diaminobenzoic acid

Table 10 Polarizability, first-order hyperpolarizability of 3ASA (monomer and dimer) computed using DFT/6-311G(d,p) method in the gas phase

Polarizability			First-order hyperpolarizability		
	monomer	dimer		monomer	dimer
α_{xx} (a.u.)	129.09	299.70	β_{xxx} (a.u)	337.79	-1.21
α_{xy} (a.u.)	1.36	2.58	β_{xxy} (a.u)	37.15	-0.13
α_{yy} (a.u.)	111.01	217.15	β_{xyy} (a.u)	84.83	-2.75
α_{xz} (a.u.)	-0.42	0.01	β_{yyy} (a.u)	21.54	2.04
α_{zy} (a.u.)	0.04	-0.01	β_{xxz} (a.u)	25.17	35.45
α_{zz} (a.u.)	40.46	79.58	β_{xyz} (a.u)	15.44	32.35
			β_{yyz} (a.u)	16.56	43.87
			β_{xzz} (a.u)	-12.67	1.82
			β_{yzz} (a.u)	13.21	-1.24
			β_{zzz} (a.u)	5.24	7.85

Table 11 Calculated polarizability and first-order hyperpolarizability of 3ASA (monomer and dimer) and compared with some reported results

Molecule	α_{total} (a.u.)	α_{total} ($\times 10^{-24}$ e.s.u.)	β_{total} (a.u)	β_{total} ($\times 10^{-30}$ e. s.u.)	References
3ASA (monomer)	93.52	13.86	487.55	4.21	*This work*
3ASA (dimer)	198.81	29.46	179.73	0.18	*This work*
Salicylic acid	–	–	299.79	2.59	[1]
5-iodosalicylic acid	–	–	231.50	2.00	[76]
Acetylsalicylic acid	104.99	15.56	208.35	1.80	[51]
3,5-diaminobenzoic acid	98.65	14.62	491.94	4.25	[23]
4-aminobenzoic acid	–	-	42.83	0.37	[77]
Urea	42.51	6.30	42.83	0.37	[52]
Thiourea	52.56	7.79	101.86	0.88	[52]
Phenyl urea	102.50	15.19	245.39	2.12	[52]

Table 12 Second-order hyperpolarizability of monomer and dimer of 3ASA computed using DFT/6-311G(d,p) method in the gas phase

Second-order hyperpolarizability		
	monomer	dimer
γ_{xxxx} (a.u.)	38,949.40	93,698.9
γ_{xxyy} (a.u.)	3922.15	14,005.6
γ_{yyyy} (a.u.)	24,652.00	10,623.3
γ_{xxzz} (a.u.)	628.67	1408.82
γ_{yyzz} (a.u.)	423.34	1242.75
γ_{zzzz} (a.u.)	236.46	548.18
γ_{total} (a.u.)	14,682.48	27,636.92
γ_{total} (e.s.u.)	7.44×10^{-36}	14.32 ($\times10^{-36}$)

[23] and phenyl urea (Table 11) [52]. Moreover, the α_{total} value of 3ASA is significantly higher than the α_{total} values of the urea and thiourea [52]. Furthermore, the estimated value of β_{total} (4.21×10^{-30} e.s.u.) value of 3ASA is higher than the reported β_{total} values of salicylic acid [1], 5-iodosalicylic acid [76], 4-aminobenzoic acid [77] and acetylsalicylic acid [51] (Table 11). The β_{total} (4.21×10^{-30} e.s.u.) of 3ASA is very closed to β_{total} (4.25×10^{-30} e.s.u.) of 3,5-diaminobenzoic acid [23]. In NLO study, urea is generally used as a reference for comparison purposes. The β_{total} of 3ASA is ~ 11 times larger than the urea, ~ 5 times greater than thiourea and ~ 2 times more than phenyl urea [52] (Table 11). The larger value of polarizability and hyperpolarizability occur due to the charge transfer within the molecular system [24]. Thus, 3ASA may have substantial applications in NLO materials.

Furthermore, the polarizability (α) and first-order hyperpolarizability (β) for dimer were also computed using the same method (DFT/B3LYP/6-311G(d,p) method) in the gas phase and listed in Tables 10 and 11. The α_{total} and β_{total} values for dimer were found to be 29.46×10^{-24} e.s.u. and 0.18×10^{-30} e.s.u., respectively. The β_{total} value of the dimer is significantly smaller than the monomer. Moreover, the β_{total} value of dimer is smaller than the urea (0.37×10^{-30} e.s.u.) [52]. These changes in results may be obtained due to the formation of dimerization through hydrogen bonding.

Additionally, the second-order hyperpolarizability (γ) for both the monomer and dimer forms of 3ASA in the gas phase has been theoretically calculated using DFT/B3LYP level of theory with a 6-311G(d,p) basis set. The calculated values of γ for the monomer and dimer are provided in Table 12. Specifically, the γ_{total} value for the monomer was computed to be 7.44×10^{-36} e.s.u., while for the dimer, it was estimated to be 14.32×10^{-36} e.s.u. This indicates that the γ_{total} value is approximately twice as high for the dimer compared to the monomer, possibly attributed to dimer formation. These findings strongly suggest the potential of 3ASA for applications in nonlinear optical (NLO) fields.

Table 13 Thermodynamic parameters of 3ASA at 298.15 K along with global minimum energy using DFT/B3LYP/6-311G(d,p)

Thermodynamic parameters	Values
SCF energy (a.u.)	-551.57
Zero-point vibrational energy (kcal mol^{-1})	86.01
Total Thermal energy, E_{th} (kcal mol^{-1})	91.81
Entropy, S (cal mol^{-1} K^{-1})	93.27
Heat capacity at const. volume, C_v (cal mol^{-1} K^{-1})	36.67
Dipole moment (Debye, D)	
μ_x	-0.57
μ_y	0.90
μ_z	0.95
μ_{total}	1.43

Table 14 Thermodynamic properties of 3ASA (rotamer P) at different temperatures calculated by DFT/B3LYP/6-311G(d,p) method

Temperature (K)	E_{th} (Kcal/mol)	S (cal/mol-K)	C_p (cal/mol-K)
50	86.34	57.71	10.27
100	86.87	66.24	15.09
200	88.77	80.27	27.04
298	91.81	93.27	38.66
400	96.09	106.14	49.09
500	101.23	118.02	57.42
600	107.23	129.10	64.03
700	113.59	139.38	69.30
800	120.59	148.92	73.58
900	127.89	157.80	77.11
1000	133.63	168.38	78.20

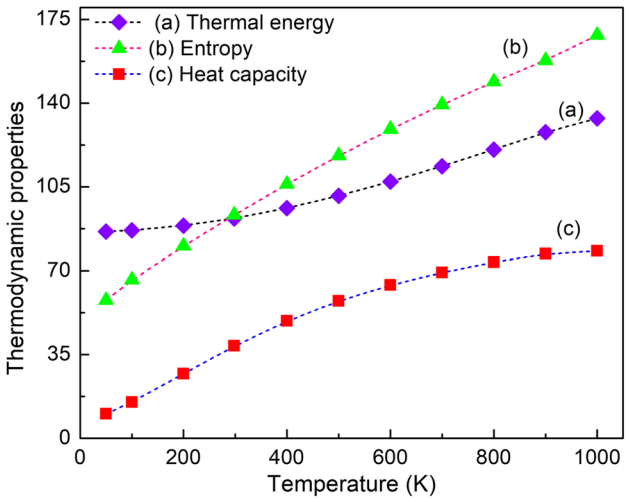


Fig. 9 Correlation graph of thermodynamic properties at different temperatures for 3ASA (rotamer P)

Thermodynamic Properties of 3ASA

The thermodynamic parameters such as zero-point vibrational energy, SCF energy, heat capacity at constant volume, thermal energy, dipole moment, and entropy, along with global minimum energy at temperature 298.15 K and pressure 1 atm have been computed employing DFT/B3LYP/6-311G(d,p) for 3ASA and are collected in Table 13.

Thermodynamics is considerably important in chemical engineering. It plays a significant role in the advance of the chemical industry, enabling design of new technologies, and avoiding challenging measurements [78]. In addition, the heat capacity is a fundamental thermodynamic parameter of a substance, and using the heat capacity information the associated thermodynamic functions such as enthalpy and entropy can be obtained using corresponding thermodynamic equations. These thermodynamic parameters are essential in examining phase change materials for energy storage applications [79].

To demonstrate the temperature dependency on the thermodynamic properties of 3ASA, heat capacity at constant pressure (C_p), total thermal energy (E_{th}) and entropy at different temperatures have been calculated by the same approach and are reported in Table 14. The temperature dependence of S, C_p and E_{th} for 3ASA is shown in Fig. 9 along with polynomial (quadratic) fitting.

From Fig. 9 and Table 14, it is noted that S, C_p, and E_{th} for 3ASA (monomer) increase on increasing the temperature from 50 to 1000 K because the molecular vibrational intensities increase with temperature [1]. Quadratic formulas were used for fitting the obtained thermodynamic parameters vs. temperature, and the resulting fitted regression parameters (R^2_{adj}) are all equal to 0.99. Therefore, the corresponding fitting equations for 3ASA are as follows,

$$S = 51.15 + 1.51 \times 10^{-2}T - 3.45 \times 10^{-5}T^2, R^2_{adj} = 0.99, \quad (12)$$

$$C_p = 2.16 + 1.41 \times 10^{-2}T - 6.66 \times 10^{-5}T^2, R^2_{adj} = 0.99, \text{and} \quad (13)$$

$$E_{th} = 84.57 + 1.72 \times 10^{-2}T + 3.33 \times 10^{-5}T^2, R^2_{adj} = 0.99 \quad (14)$$

These thermodynamic parameters for 3ASA are helpful to ascertain the direction of chemical reactions and to further estimate thermodynamic properties dealing with thermodynamic functions and the second law of thermodynamics [80].

Conclusion

The molecular structures of 3ASA (monomer and dimer) were optimized using DFT and TD-DFT theories in the ground and excited states and significant structural, molecular and thermodynamic parameters were obtained. The small energy gap suggests that both monomer and dimer are reactive. MEP map shows that the carbonyl oxygen is in a position of electrophilic attack in the monomer. Knowledge of MEP map is useful in understanding and foreseeing molecular interactions. Depending on the medium, the calculated absorption and emission maxima of 3ASA were observed in the range of 343—347 nm and 429 – 448 nm, respectively. The theoretical absorption and emission spectra in most of the solvents are closer to the experimental results. Moreover, the computed absorption and emission spectra of monomer and dimer are in correlation with each other. The radiative lifetime, light harvesting efficiency (LHE) and band gap energy show that 3ASA may be useful in photovoltaics and wide bandgap power devices. Moreover, the NBO analysis shows that charge transfer interactions contribute significantly in stabilizing the molecular system. In the case of dimer, H-bonding plays a dominant role in stabilizing the molecular framework. Furthermore, the calculated NLO parameters indicate that 3ASA can be used in NLO applications. The thermodynamic properties: entropy, total thermal energy and heat capacity at constant pressure increase with the increase in temperature as a result of rising molecular vibrational intensities. Thus, the thermodynamic study is important to predict the direction of chemical reactions and other thermodynamic parameters. Hence, the present study delivers detailed structural/molecular information of the 3ASA molecule.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10895-024-03781-y.

Acknowledgements The authors are grateful to Prof. N. G. Sahoo, Department of Chemistry, DSB campus, Kumaun University, Nainital, for UV–vis spectrophotometer for initial measurement of absorption spectra and Dr. H. C. Joshi, Institute of Plasma Research, Bhat, Gandhinagar for valuable comments and suggestions.

Author contribution **SH**: Conceptualization, data curation, formal analysis, methodology, investigation, validation, visualization and writing—original draft. **NP**: Writing—review & editing. **NF:** Writing—review & editing. **SP**: Supervision. **MSM**: Supervision, project administration, resources, writing—review & editing.

Funding Not applicable.

Data availability All data generated or analysed during this study are included in this published article and supplementary information.

Declarations

Competing Interest The authors declare that they have no known financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Consent to Publish All authors have approved the paper and agree with its publication.

References

1. Suresh S, Gunasekaran S, Srinivasan S (2014) Spectroscopic (FT-IR, FT-Raman, NMR and UV–Visible) and quantum chemical studies of molecular geometry, Frontier molecular orbital, NLO, NBO and thermodynamic properties of salicylic acid. Spectrochim Acta Part A Mol Biomol Spectrosc 132:130–141
2. Faize L, Faize M (2018) Functional analogues of salicylic acid and their use in crop protection. Agronomy 8(1):5
3. Sebastian S, Sylvestre S, Jayabharathi J, Ayyapan S, Amalanathan M, Oudayakumar K, Herman IA (2015) Study on conformational stability, molecular structure, vibrational spectra, NBO, TD-DFT, HOMO and LUMO analysis of 3, 5-dinitrosalicylic acid by DFT techniques. Spectrochim Acta Part A Mol Biomol Spectrosc 136:1107–1118
4. Silva AARD, Lima GSD, Azevedo CAVD, Gheyi HR, Souza ARD, Fernandes PD (2021) Salicylic acid relieves the effect of saline stress on soursop morphysiology. Ciência e Agrotecnologia, 45
5. Cremlyn RJW (1991) Agrochemicals: Preparation and Mode of Actions. John Wiley & Sons
6. Arsianti A, Tanimoto H, Morimoto T, Bahtiar A, Takeya T, Kakiuchi K (2012) Synthesis and anticancer activity of polyhydroxylated 18-membered analogue of antimycin A3. Tetrahedron 68(13):2884–2891
7. Grimes PE (2011) Salicylic acid. In Color atlas of chemical peels (pp. 17–24). Berlin, Heidelberg: Springer Berlin Heidelberg
8. Joshi HC, Mishra H, Tripathi HB (1997) Photophysics and photochemistry of salicylic acid revisited. J Photochem Photobiol, A 105(1):15–20
9. Arora P, Joshi HC, Joshi NK, Pandey N, Pandey P, Pant S (2014) Modulation of the fluorescence properties of 5-amino salicylic acid by triethylamine. J Mol Liq 191:128–136
10. Joshi HC, Antonov L (2021) Excited-state intramolecular proton transfer: A short introductory review. Molecules 26(5):1475
11. Maheshwari S, Chowdhury A, Sathyamurthy N, Mishra H, Tripathi HB, Panda M, Chandrasekhar J (1999) Ground and excited state intramolecular proton transfer in salicylic acid: An ab initio electronic structure investigation. J Phys Chem A 103(31):6257–6262
12. Suban KS, Rati KB, Bikram KK, Minati B (2009) Excited State Intramolecular Proton Transfer in Methyl and Methoxy Substituted Salicylic Acid. Chin J Chem 27(10):1897–1902
13. Paul BK, Samanta A, Guchhait N (2010) Deciphering the photophysics of 5-chlorosalicylic acid: evidence for excited-state intramolecular proton transfer. Photochem Photobiol Sci 9:57–67
14. Paul BK, Guchhait N (2012) Evidence for excited-state intramolecular proton transfer in 4-chlorosalicylic acid from combined experimental and computational studies: Quantum chemical treatment of the intramolecular hydrogen bonding interaction. Chem Phys 403:94–104
15. Paul BK, Samanta A, Guchhait N (2011) On the photophysics of 3, 5, 6-trichlorosalicylic acid: Spectroscopic study combined with Hartree-Fock and density functional theory calculations. J Fluoresc 21:1265–1279
16. Jang S, Jin SI, Park CR (2007) TDDFT potential energy functions for excited state intramolecular proton transfer of salicylic acid, 3-aminosalicylic acid, 5-aminosalicylic acid, and 5-methoxysalicylic acid. Bull Korean Chem Soc 28(12):2343–2353
17. Joshi HC, Gooijer C, van der Zwan G (2003) Substitution effects on the photophysical characteristics of the salicylic anion. J Fluoresc 13:227–234
18. Suyal K, Joshi NK, Rautela R, Joshi HC, Pant S (2010) Fluorescence properties of 4-amino salicylic acid in polymers. J Photochem Photobiol, A 216(1):51–58
19. Akkaya Y, Akyuz S (2006) Infrared and Raman spectra, ab initio calculations vibrational assignment of 4-aminosalicylic acid. Vib Spectrosc 42(2):292–301
20. Singh YP, Das R, Singh RA (2007) Numerical simulation of the internal vibrations of COOH group in amino-salicylic acids. Afr J Biochem Res 1(2):19–23
21. Moradi M, Rahimpour E, Hemmati S, Kuentz M, Martinez F, Jouyban A (2020) Solubility of 3-aminosalicylic acid in ethanol+ water mixtures at different temperatures. J Mol Liq 318:114310
22. Nogueira HI (1998) Surface-enhanced Raman scattering (SERS) of 3-aminosalicylic and 2-mercaptonicotinic acids in silver colloids. Spectrochim Acta Part A Mol Biomol Spectrosc 54(10):1461–1470
23. Husain S, Pandey N, Fatma N, Pant S, Mehata MS (2022) Spectral characteristics of 3, 5-diaminobenzoic acid in pure and mixed solvents: Experimental and theoretical study. J Mol Liq 368:120783
24. Pandey N, Mehata MS, Pant S, Tewari N (2021) Structural, Electronic and NLO properties of 6-aminoquinoline: A DFT/TD-DFT study. J Fluoresc 31(6):1719–1729
25. Bocharnikova EN, Tchaikovskaya ON, Bazyl OK, Artyukhov VY, Mayer GV (2020) Theoretical study of bisphenol A photolysis. In Adv Quant Chem (Vol. 81, pp. 191–217). Academic Press
26. Karelson M, Lobanov VS, Katritzky AR (1996) Quantum-chemical descriptors in QSAR/QSPR studies. Chem Rev 96(3):1027–1044
27. Janjua MRSA, Su ZM, Guan W, Liu CG, Yan LK, Song P, Maheen G (2010) Tuning second-order non-linear (NLO) optical response of organoimido-substituted hexamolybdates through halogens: Quantum design of novel organic-inorganic hybrid NLO materials. Aust J Chem 63(5):836–844
28. Janjua MRSA, Guan W, Yan L, Su ZM, Karim A, Akbar J (2010) Quantum Chemical Design for Enhanced Second-Order NLO Response of Terpyridine-Substituted Hexamolybdates
29. Abdullah MI, Janjua MRSA, Nazar MF, Mahmood A (2013) Quantum chemical designing of efficient TC4-based sensitizers by modification of auxiliary donor and π-spacer. Bull Chem Soc Jpn 86(11):1272–1281
30. Irfan A, Hussien M, Mehboob MY, Ahmad A, Janjua MRSA (2022) Learning from Fullerenes and Predicting for Y6: Machine Learning and High-Throughput Screening of Small Molecule Donors for Organic Solar Cells. Energ Technol 10(6):2101096
31. Janjua MRSA (2022) Photovoltaic properties and enhancement in near-infrared light absorption capabilities of acceptor materials for organic solar cell applications: A quantum chemical perspective via DFT. J Phys Chem Solids 171:110996
32. Janjua MRSA (2021) Prediction and understanding: Quantum chemical framework of transition metals enclosed in a B12N12 inorganic nanocluster for adsorption and removal of DDT from the environment. Inorg Chem 60(14):10837–10847
33. Janjua MRSA, Jamil S, Mahmood A, Zafar A, Haroon M, Bhatti HN (2015) Solvent-dependent non-linear optical properties of 5, 5′-disubstituted-2, 2′-bipyridine complexes of ruthenium (II): a quantum chemical perspective. Aust J Chem 68(10):1502–1507

34. Janjua MRSA (2021) Quantum chemical design of D–π–A-type donor materials for highly efficient, photostable, and vacuum-processed organic solar cells. Energ Technol 9(10):2100489
35. Janjua MRSA (2021) Theoretical framework for encapsulation of inorganic B12N12 nanoclusters with alkaline earth metals for efficient hydrogen adsorption: a step forward toward hydrogen storage materials. Inorg Chem 60(4):2816–2828
36. Janjua MRSA, Mahmood A, Ahmad F (2013) Solvent effects on nonlinear optical response of certain tetramminеruthenium (II) complexes of modified 1, 10-phenanthrolines. Can J Chem 91(12):1303–1309
37. Haroon M, Janjua MRSA (2022) Exploring the effect of end-capped modifications of carbazole-based fullerene-free acceptor molecules for high-performance indoor organic solar cell applications. J Comput Electron, 1–12
38. Mahmood R, Janjua MRSA, Jamil S (2017) DFT molecular simulation for design and effect of core bridging acceptors (BA) on NLO response: first theoretical framework to enhance nonlinearity through BA. J Cluster Sci 28:3175–3183
39. Janjua MRSA (2017) First theoretical framework of di-substituted donor moieties of triphenylamine and carbazole for NLO properties: quantum paradigms of interactive molecular computation. Mol Simul 43(18):1539–1545
40. Ghosh A (2006) Just how good is DFT? J Biol Inorg Chem 11(6):671–673
41. Mehata MS, Singh AK, Sinha RK (2016) Experimental and theoretical study of hydroxyquinolines: hydroxyl group position dependent dipole moment and charge-separation in the photoexcited state leading to fluorescence. Meth Appl Fluoresc 4(4):045004
42. Govindarajan M, Karabacak M (2012) Spectroscopic properties, NLO, HOMO–LUMO and NBO analysis of 2, 5-Lutidine. Spectrochim Acta Part A Mol Biomol Spectrosc 96:421–435
43. Li L, Cai T, Wang Z, Zhou Z, Geng Y, Sun T (2014) Study on molecular structure, spectroscopic investigation (IR, Raman and NMR), vibrational assignments and HOMO–LUMO analysis of L-sodium folinate using DFT: A combined experimental and quantum chemical approach. Spectrochim Acta Part A Mol Biomol Spectrosc 120:106–118
44. Alam MJ, Ahmad S (2015) FTIR, FT-Raman, UV–visible spectra and quantum chemical calculations of allantoin molecule and its hydrogen bonded dimers. Spectrochim Acta Part A Mol Biomol Spectrosc 136:961–978
45. Karabacak M, Sinha L, Prasad O, Cinar Z, Cinar M (2012) The spectroscopic (FT-Raman, FT-IR, UV and NMR), molecular electrostatic potential, polarizability and hyperpolarizability, NBO and HOMO–LUMO analysis of monomeric and dimeric structures of 4-chloro-3, 5-dinitrobenzoic acid. Spectrochim Acta Part A Mol Biomol Spectrosc 93:33–46
46. Al-Ahmary KM, Habeeb MM, Aljahdali SH (2019) Synthesis, spectroscopic studies and DFT/TD-DFT/PCM calculations of molecular structure, spectroscopic characterization and NBO of charge transfer complex between 5-amino-1, 3-dimethylpyrazole (5-ADMP) with chloranilic acid (CLA) in different solvents. J Mol Liq 277:453–470
47. Amalanathan M, Rastogi VK, Joe IH, Palafox MA, Tomar R (2011) Density functional theory calculations and vibrational spectral analysis of 3, 5-(dinitrobenzoic acid). Spectrochim Acta Part A Mol Biomol Spectrosc 78(5):1437–1444
48. Mathammal R, Sangeetha K, Sangeetha M, Mekala R, Gadheeja S (2016) Molecular structure, vibrational, UV, NMR, HOMO-LUMO, MEP, NLO, NBO analysis of 3, 5 di tert butyl 4 hydroxy benzoic acid. J Mol Struct 1120:1–14
49. Karabacak M, Cinar Z, Kurt M, Sudha S, Sundaraganesan N (2012) FT-IR, FT-Raman, NMR and UV–vis spectra, vibrational assignments and DFT calculations of 4-butyl benzoic acid. Spectrochim Acta Part A Mol Biomol Spectrosc 85(1):179–189
50. Glory DCM, Madivanane R, Sambathkumar K (2015) Electronic Structure Investigations of 3 and 5-Diamino-1, 2, 4-Triazole By UV–Visible, NMR spectral studies and homo-lumo analysis by AB Initio and DFT calculations. Elixir Comp Chem 89:36730–36741
51. Rana M, Yadav P, Chowdhury P (2023) DFT investigation of nonlinear optical response of organic compound: acetylsalicylic acid. Intl J Mater Res, (0)
52. Lakhera S, Rana M, Devlal K, Sharma S, Chowdhury P, Dhuliya V ... Girisun TS (2023) Exploring the nonlinear optical limiting activity of para-aminobenzoic acid by experimental and DFT approach. J Photochem Photobiol A: Chem, 114987
53. Janjua MRSA, Yamani ZH, Jamil S, Mahmood A, Ahmad I, Haroon M, ... Pan S (2015) First principle study of electronic and non-linear optical (NLO) properties of triphenylamine dyes: interactive design computation of new NLO compounds. Austr J Chem, 69(4), 467–472
54. Dennington R, Keith T, Millam J (2009) GaussView, version 5
55. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, BrothersE, Kudin KN, Staroverov VN, Keith T, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2010) Gaussian 09, Revision B.01, Gaussian, Inc., Wallingford CT
56. Tirado-Rives J, Jorgensen WL (2008) Performance of B3LYP density functional methods for a large set of organic molecules. J Chem Theory Comput 4(2):297–306
57. Stratmann RE, Scuseria GE, Frisch MJ (1998) An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules. J Chem Phys 109(19):8218–8224
58. Cances E, Mennucci B, Tomasi J (1997) A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. J Chem Phys 107(8):3032–3041
59. Skyner RE, McDonagh JL, Groom CR, van Mourik TV, Mitchell JBO (2015) A review of methods for the calculation of solution free energies and the modelling of systems in solution. Phys Chem Chem Phys 17(9):6174–6191
60. Husain S, Mehata MS, Pandey N, Mishra H, Pant S (2021) Reinvestigation of the photophysics of 3-aminobenzoic acid in neat and mixed binary solvents. Spectrochim Acta Part A Mol Biomol Spectrosc 247:119100
61. Fatma N, Mehata MS, Pandey N, Pant S (2021) Experimental and theoretical interpretations of spectral behavior of 6-methoxyflavone. J Photochem Photobiol, A 404:112945
62. Srivastava K, Shimpi MR, Srivastava A, Tandon P, Sinha K, Velaga SP (2016) Vibrational analysis and chemical activity of paracetamol–oxalic acid cocrystal based on monomer and dimer calculations: DFT and AIM approach. RSC Adv 6(12):10024–10037
63. Stalin T, Rajendiran N (2006) Intramolecular charge transfer associated with hydrogen bonding effects on 2-aminobenzoic acid. J Photochem Photobiol, A 182(2):137–150

64. Stalin T, Shanthi B, Rani PV, Rajendiran N (2006) Solvatochromism, prototropism and complexation of para-aminobenzoic acid. J Incl Phenom Macrocycl Chem 55:21–29
65. Tiwari K, Arora P, Pandey N, Pandey P, Joshi HC, Pant S (2014) Experimental and computational approaches on dipole moment of 5-aminoisoquinoline. J Mol Liq 200:460–464
66. Manohara SR, Kumar VU, Gerward L (2013) Estimation of ground and excited-state dipole moments of 1, 2-diazines by solvatochromic method and quantum-chemical calculation. J Mol Liq 181:97–104
67. Bibi S, Farooq F, Bai FQ, Zhang HX (2021) DFT and TD-DFT studies of phenothiazine based derivatives as fluorescent materials for semiconductor applications. Mater Sci Semicond Process 134:106036
68. Rohatgi-Mukherjee KK (1978) Fundamentals of photochemistry. New Age International
69. Calogero G, Yum JH, Sinopoli A, Di Marco G, Grätzel M, Nazeeruddin MK (2012) Anthocyanins and betalains as light-harvesting pigments for dye-sensitized solar cells. Sol Energy 86(5):1563–1575
70. Weatherby K, Carter D (2013) Chromera velia: the missing link in the evolution of parasitism. In Adv Appl Microbiol (Vol. 85, pp. 119–144). Academic Press
71. Pandey N, Tewari N, Pant S, Mehata MS (2022) Solvatochromism and estimation of ground and excited state dipole moments of 6-aminoquinoline. Spectrochim Acta Part A Mol Biomol Spectrosc 267:120498
72. Millán J (2007) Wide band-gap power semiconductor devices. IET Circuits Devices Syst 1(5):372
73. Kumar A, Moradpour M, Losito M, Franke WT, Ramasamy S, Baccoli R, Gatto G (2022) Wide band gap devices and their application in power electronics. Energies 15(23):9172
74. Karakas A, Ceylan Y, Karakaya M, Taser M, Terlemez BB, Eren N ... Sahraoui B (2019) Theoretical diagnostics of second and third-order hyperpolarizabilities of several acid derivatives. Open Chem, 17(1), 151–156
75. Prakasam M, Anbarasan PM (2016) Second order hyperpolarizability of triphenylamine based organic sensitizers: a first principle theoretical study. RSC Adv 6:75242–75250
76. Karaca C, Atac A, Karabacak M (2015) Conformational analysis, spectroscopic study (FT-IR, FT-Raman, UV, 1H and 13C NMR), molecular orbital energy and NLO properties of 5-iodosalicylic acid. Spectrochim Acta Part A Mol Biomol Spectrosc 136:295–305
77. Borah B, Devi TG (2020) Vibrational study on the molecular interaction of L-Proline and Para-Aminobenzoic acid. J Mol Struct 1203:127396
78. Popovic M (2019) Thermodynamic properties of microorganisms: determination and analysis of enthalpy, entropy, and Gibbs free energy of biomass, cells and colonies of 32 microorganism species. Heliyon 5(6):e01950
79. Kou Y, Wang S, Luo J, Sun K, Zhang J, Tan Z, Shi Q (2019) Thermal analysis and heat capacity study of polyethylene glycol (PEG) phase change materials for thermal energy storage applications. J Chem Thermodyn 128:259–274
80. Nataraj A, Balachandran V, Karthick T (2013) Molecular structure, vibrational spectra, first hyperpolarizability and HOMO–LUMO analysis of p-acetylbenzonitrile using quantum chemical calculation. J Mol Struct 1038:134–144

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

Quantum Resistant Blockchain-based Architecture for Secure Medical Data Sharing

Kanupriya Jain
Department of CSE
Delhi Technological University
Delhi, India
kanupriyajain12@gmail.com

Mahima Singh
Department of CSE
Delhi Technological University
Delhi, India
mahimasinghams@gmail.com

Harshita Gupta
Department of CSE
Delhi Technological University
Delhi, India
gupta.harshita0512@gmail.com

Aruna Bhat
Department of CSE
Delhi Technological University
Delhi, India
aruna.bhat@dtu.ac.in

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575286

***Abstract*— The development of quantum computing substantially threatens modern cryptographic functions in blockchains, making it unsafe to share sensitive details such as Electronic Medical Records (EMRs). To address this threat, the paper proposes a novel and secure architecture for EMR data sharing using permissioned blockchain and quantum-resistant cryptography. Blockchain technology is used to provide data immutability, transparency, and integrity. Hybrid encryption with CRYSTALS-Kyber, a post-quantum cryptographic algorithm for key encapsulation and AES-256 for symmetric data encryption is used to address the security concerns in conventional blockchains. The patient stores their access policies that control who can access their data. The data access is managed via the patient-controlled re-encryption mechanism through which the data requester may safely access the EMR. The paper presents a detailed analysis of the system's security, patient privacy, and data ownership control, offering a patient-centric solution for secure EMR sharing in the coming quantum age.**

Keywords*— *Blockchain, Post-Quantum, CRYSTALS-Kyber, Electronic Medical Record (EMR), Cryptography

I. INTRODUCTION

In modern healthcare facilities, Electronic Medical Records (EMRs) are electronic versions of a patient's medical data maintained by healthcare providers that contain their relevant medical information [1]. These are usually used to store patients' private data that can be useful in diagnosing, treating or researching. However, this data should be shared safely among authorized individuals, including lab technicians, doctors, and the patients themselves, so that they can receive the right medical care [1]. The traditional systems prevalent today make it difficult to balance between ease of access, data privacy and security. This, in turn, complicates the implementation of access control while ensuring data confidentiality as well as security since there are diverse parties affected by this problem. Moreover, there is a lack of accountability or confidence in the central authority leading to loss of data during storage and transmission among interested stakeholders [2], [3].

The secure sharing of EMRs while ensuring transparency, immutability and decentralization is possible through blockchain-based EMR systems. Being a decentralized approach, blockchain eliminates security concerns and unauthorized access to data. Its immutability ensures that data intractability and integrity [4]. Nonetheless, the threat of potential quantum computing attacks on prevailing cryptographic methods calls into question the long-term security of Blockchain-based EMR Systems. Quantum computers are equipped with certain properties such as superposition, entanglement and interference that enable them to resolve specific problems much faster than classical computers using specialized algorithms for those purposes [5]. Technological advancements necessitate the shift towards PQC approaches in securing the existing systems. The motivation lies in the need to build systems incorporating PQC algorithms.

A. Problem Statement

Quantum computing is a rapidly emerging field that utilizes quantum mechanics principles to perform information processing. This makes the industry very interesting for many sectors because quantum computing has well-established theoretical basis as well as experimental proof. However, if this new paradigm ever becomes a reality with sufficient qubits, it could spell doom for blockchain systems, which rely on pre-quantum cryptographic methods and are hence susceptible to attack from such computers. Shor's algorithm [6] among other quantum algorithms, can effectively break cryptographic problems at the heart of blockchain, like integer factorization, discrete logarithmic problems or elliptic curve intractability in polynomial time, which were computationally inefficient for classical computers. The Grover's Algorithm [7] in quantum computing can efficiently produce hashes and change any block without altering data integrity.

These advancements, especially the claim of achieving quantum supremacy, risk the confidentiality and integrity of traditional public key cryptosystems used in implementing blockchains. Hence, there is a pressing need for blockchain networks to be resistant to future quantum attacks. They need to be redesigned with the help of quantum-resistant cryptosystems, thus developing post-quantum blockchain cryptosystems. Lightweight and computational efficiency are required for such architectures [8]. More research is needed to overcome these vulnerabilities, specifically for sensitive medical data.

EMRs act as the centralized data source for a patient's sensitive medical data like demographics, medications, lab results, etc. This medical data is crucial for better healthcare quality, reduced costs and secure information sharing among authorized providers. Hence, to secure medical data in today's healthcare systems, blockchain-based EMR systems need to be secured and seamlessly incorporate efficient PQC algorithms into existing blockchain infrastructures.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

B. Objectives and Contributions

The research aims to address the upcoming challenge of the necessity of quantum-resistant blockchain systems for EMRs and hence demonstrates the potential of utilizing PQC techniques to secure EMRs within a blockchain system for enhanced data security. The key objectives and contributions of the paper are as follows:

- To propose a novel blockchain-based architecture for secure medical data sharing among healthcare providers, patients, and authorized data requestors resistant to potential attacks from quantum computers.
- To utilize a hybrid encryption approach combining CRYSTALS-Kyber, a quantum-resistant key encapsulation mechanism, and AES-256, offering enhanced security for EMR data.
- The architecture is designed to prioritize patient control over data by allowing them to define access policies and facilitates secure data retrieval using re-encryption for authorized data requestors.

The organization of the paper is as follows. The related work summarizing the existing literature is discussed in Section 2. Section 3 overviews the technical background of blockchain and quantum-resistant cryptography. Section 4 proposes the novel quantum-resistant blockchain architecture and its components. The implementation, workflow, and algorithmic details are explained in Section 5. Section 6 discusses the security analysis and section 7 concludes the paper.

II. Related Works

In [2], a quantum blockchain approach for EMR privacy is proposed, which utilizes quantum communication, entanglement, and quantum key distribution protocols. Performance analysis shows that the proposed system can overcome intercept, entangle-measure, man-in-the-middle, intercept-resend, collective and coherent attacks. But it's difficult to determine the system's efficacy as no detailed implementation is given. In [9] a blockchain-based Electronic Health Records (EHR) system, implemented on Hyperledger Fabric, with user access control algorithms and quantum blind signatures for key distribution is presented. Although it enhances the security, it does not provide a detailed analysis of potential weaknesses and there are scalability challenges. [10] discusses the advantages of combining blockchain and quantum technologies in healthcare but does not give specific details of how quantum blockchain can be implemented in healthcare systems.

Apart from healthcare systems, research has also been done on other blockchain systems to make them quantum-resistant. In [11], a conceptual quantum-resistant blockchain system based on the Linux Foundation Hyperledger Fabric, for university credential verification is proposed. It utilizes Lattice-Based Cryptosystems along with Leighton-Micali Signature (LMS) (Hash-Based Signatures) and eXtended Merkle Signature Scheme (XMSS) for quantum-resistant public key cryptography and hashing respectively. Benchmark testing is needed to address performance uncertainties. [12] introduces a quantum-resistant e-voting protocol on blockchain combining code-based Niederreiter algorithm. Its security is based on the syndrome decoding problem of coding theory and a certificateless traceable ring signature algorithm. But it is efficient only for small-scale elections.

In [13] a quantum-resistant blockchain scheme is proposed using qTESLA, a digital signature based on lattice cipher for smart city applications. It stores public keys and signatures on the IPFS network to increase transaction speed and system performance. However, it doesn't allow parallel transactions. [14] explores PQC implementation in Federated Learning (FL) systems and proposes a hybrid approach combining Dilithium/Falcon with XMSS for security in blockchain-based FL systems. They utilized verifiable random functions to enhance security and performance. However, PQC schemes are still in the standardization phase creating concerns regarding their suitability for various applications and hardware systems.

[15] presents a blockchain voting system that utilizes Practical Byzantine Fault Tolerance for consensus, stateful hash-based signatures (LMS/XMSSMT/HSS), lattice-based and Super Singular Elliptic Curve Isogeny-based cryptic algorithms for public key encryption, but complete automation and device-level security are still concerns. In [16], a framework is introduced to secure the quantum-resistant cloud-based blockchain storage. It utilized a combination of Quantum key distribution (QKD) for key generation, CRYSTALS-Kyber for data encryption, and Zero-Knowledge Proof (ZKP) for user verification before giving access to decryption keys. However, further development is still needed to enhance its scalability and compatibility with existing blockchain systems.

In summary, there is previous research on quantum-resistant cryptography, but the work lacks focus on integrating PQC algorithms in blockchain-based EMR sharing or healthcare systems to enhance their security. The existing papers also lack post-quantum Key Encapsulation Mechanisms, and NIST-selected algorithms in their blockchain systems as they are relatively new. Recent research also failed to address the weaknesses in data sharing, storage, and access in these blockchain systems.

III. BACKGROUND

This section provides an introduction to Blockchain and PQC's technical background.

A. Blockchain

Blockchain is an immutable, decentralized, and shared ledger of transactions that can safely store asset information, eliminating the role of trusted intermediaries [17], [18]. Its base lies in cryptocurrency and its first appearance was when Satoshi Nakamoto introduced it as a peer-to-peer cash system [19]. A blockchain is formed by joining many blocks chronologically where each block is made up of a header and a body [20]. The header stores the previous block's hash, timestamps, and Merkle Root. In turn, the body contains the actual data that varies based on the type of blockchain. Consensus algorithms like Proof of Work (PoW) and Proof of Stake (PoS) are used to approve every transaction [21]. It determines which nodes in the network have the right to record requested transactions and allows them to validate them efficiently. Blockchain decentralization makes it tamper-proof and immune to any malicious alterations in

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

data. Blockchains use smart contracts to establish trust. They are predefined rules which are triggered automatically when certain terms and conditions, established beforehand, are satisfied. Hence, blockchain technology is the best solution for trustless environments like finance, voting systems, education systems, and healthcare systems [22].

B. Quantum Resistant Cryptography

The pre-quantum cryptography's security is based upon the hardness of mathematical problems like factorization of integers, discrete logarithm problems, and intractability of elliptic curves [23]. PQC is comprised of various algorithms that are built on different complex problems to withstand quantum attacks including cryptography that is code-based, hash-based, lattice-based and multi-variate polynomial cryptography. Lattice-based cryptography is based on hard problems defined on lattices such as the Shortest Vector Problem (SVP) or the Learning with Errors (LWE) problem [24]. Multivariate polynomial cryptography is based on the difficulty of solving systems of multivariate polynomial equations over a finite field [25]. Code-based cryptography is built on error-correcting codes like McEliece cryptosystems or Goppa codes [26], whereas Hash-based Cryptography uses special functions called cryptographic hash functions for digital signatures [27].

The threats of advanced quantum computers in 2016 made the National Institute of Standards and Technology (NIST) begin identifying and standardizing algorithms for PQC [28]. By 2017, it is stated that there were 82 algorithms submitted in total and only 69 candidates managed to be selected to represent the first round. The second-round evaluation was conducted in 2019 with 26 algorithms chosen by NIST and by the year 2020, seven finalists and eight alternates were shortlisted. In 2022, after several rounds of shortlisting and rigorous evaluation, 4 PQC algorithms were selected for standardization [29]. These are CRYSTALS-Kyber [30], a public key encapsulation mechanism along with three digital signature schemes: CRYSTALS-Dilithium [31], Falcon [32], and SPHINCS+ [33]. SPHINCS+ utilizes hash functions for security while the other three are based on lattice cryptography.

C. CRYSTALS-Kyber

CRYSTALS-Kyber is a member of CRYSTALS (Cryptographic Suite for Algebraic Lattices) standardized by NIST for PQC [34]. Kyber is an IND-CCA2 (Indistinguishability under adaptive Chosen Ciphertext Attack) secure KEM [30]. It is a cryptosystem that is lattice-based with security based upon the hardness of solving a variant of LWE problem over module lattices. The CRYSTALS-Kyber structure consists of two components KYBER.CCAKEM and KYBER.CPAPKE. The KYBER.CCAKEM is a KEM algorithm resistant to the most effective security assurance for public key encryption schemes like IND-CCA2 and KYBER.CPAPKE is an asymmetric encryption scheme that can withstand IND-CPA [35]. The KEM encapsulates bit strings (keys) which are randomly selected to produce a randomly chosen key and encapsulated key. This key can encrypt the data using symmetric encryption [32]. The key encapsulation mechanism involves three processes: key generation, encapsulation, and decapsulation [36]. Kyber-1024, Kyber-768, and Kyber-512 are three variants that differ in coefficient distribution and module dimensions and have security equivalent to AES-256, AES-192, and AES-128 respectively [36].

To create the Kyber key-pair after selecting system parameters, the private key is composed of small coefficient polynomials, while the public key is formed by matrix multiplication of a random matrix with the private key vector and the addition of a random error vector. This process is illustrated in Fig. 1. Kyber encapsulation then uses the public key and randomness via small coefficient polynomials to create a ciphertext that hides a secret shared key generated during the process. Kyber decapsulation uses the private key polynomials to remove the randomness embedded in the ciphertext using polynomial operations to recover the secret shared key. The randomness provides post-quantum security.

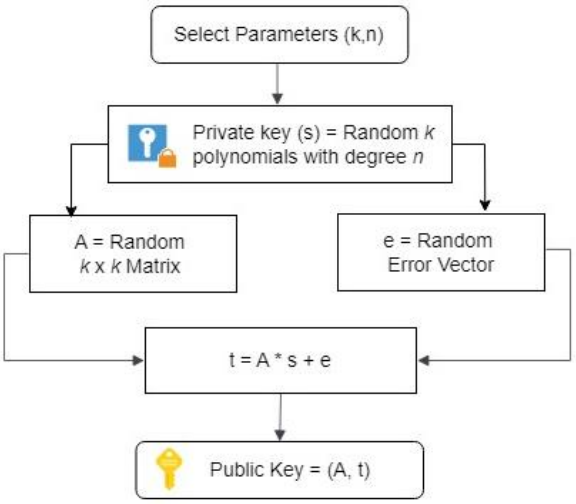


Fig. 1 Generation of Public and Private Key using CRYSTALS-Kyber

IV. PROPOSED MODEL

This section describes the system's overall components, the blockchain's detailed block structure, the hybrid data storage approach, and the cryptographic components used in the proposed quantum-resistant blockchain architecture.

A. System Components

The proposed architecture for secure sharing of EMR data among different stakeholders consists of the following system components:

- Users: These represent the different clients interacting with the system, like Patients, Healthcare Professionals, and Data Requestors.
- Certificate Authority: This supervising authority issues certificates to users after verifying their public keys and other identification documents.
- IPFS (Interplanetary File Storage): It is a decentralized, cost-effective and secure storage for encrypted EMRs.
- Blockchain Network: A permissioned blockchain network is used, as a digital ledger technology for secure data sharing and recording.

B. Blockchain Structure

A permissioned blockchain is used to ensure controlled access and authorization of participants. Each block consists

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

of a header and a body [37]. Fig. 2 represents the structure of a block and a transaction within this block. The block header contains essential data for maintaining blockchain integrity, including: a version of the blockchain protocol used, a hash of the current block, a hash of the previous block to link the blocks chronologically, a nonce that is used in proof-of-work mechanism to generate a valid hash of the block, a Merkle root representing all the transactions in the body to facilitate the verification of transaction data within the block and a timestamp. The body of the block stores the actual transaction data as a list of transactions authenticated with digital signatures.

Each transaction contains critical information about the EMR data being shared, including the transaction ID, patient ID, type of EMR data, reference ID of the EMR uploaded on IPFS, a hash of the encrypted EMR data to ensure data integrity, the patient's access policy specifying his access permissions to ensure authorized access, the quantum-resistant cipher key used for encryption, and other relevant metadata.

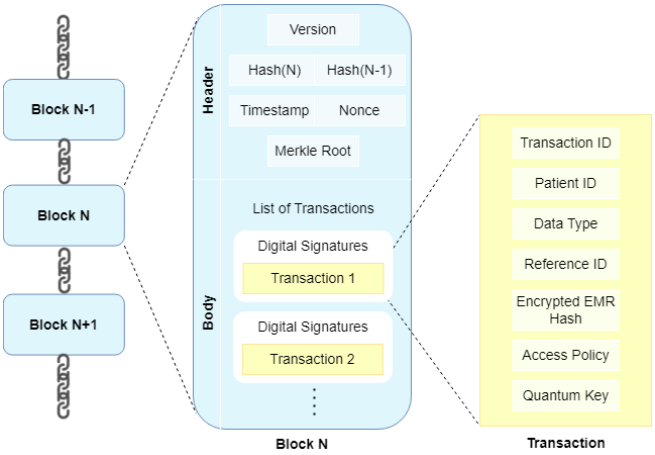


Fig. 2 Structure of a Block

C. Data Storage

A hybrid data storage mechanism is used that combines the strengths of blockchain and off-chain storage. The actual EMR data is encrypted with a quantum-resistant algorithm and stored securely in an off-chain decentralized database. The relevant metadata, such as the hash of the encrypted EMR, is stored on the blockchain. Since every node in the blockchain network has a copy of the ledger, this hybrid approach for data storage reduces storage costs, improves transaction processing speed and enhances the system's scalability.

IPFS (Interplanetary File Storage) [38] is used for the off-chain database. It is a decentralized storage designed to store encrypted data across multiple nodes when the size of the data is more than a threshold (size > 256KB), eliminating the role of a central server [39]. It stores the encrypted EMR and returns an IPFS Content Identifier for data retrieval. Due to its decentralized nature, IPFS is advantageous over the present cloud storage solutions. It eliminates a single point of failure, translates to high storage throughput, and improves data retrieval speeds [39].

D. Cryptographic Components

A hybrid encryption mechanism is used to securely encrypt the EMR, consisting of a post-quantum key encapsulation mechanism: CRYSTALS-Kyber and a symmetric encryption algorithm: AES-256. This is illustrated in Fig. 3. The key is encapsulated using the CRYSTALS-Kyber algorithm and the transaction data is encrypted using the AES-256 algorithm. AES is a block cipher that uses substitution, shifting, mixing, and key addition operations to achieve symmetric encryption [40]. AES-256 provides efficient encryption with a strong 256-bit key length, making it a suitable choice [41]. Combining asymmetric primitives for key establishment and symmetric cryptography for data encryption reduces the computational overhead [42].

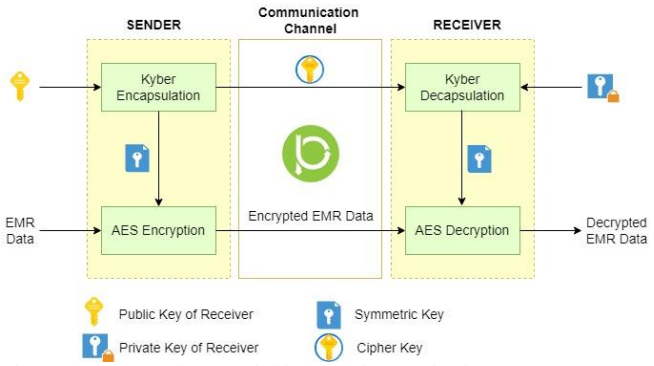


Fig. 3 Quantum-Resistant Hybrid Encryption Mechanism

First, Kyber encryption generates a key-pair comprising a symmetric and cipher key. The cipher key is the symmetric key encapsulated with the receiver's public key. The sender encrypts the EMR data with the symmetric key using AES-256 and transmits the encrypted EMR data and the cipher key to the receiver. The Kyber decryption algorithm decapsulates the cipher key into the symmetric key using the receiver's private key. The decapsulated symmetric key is finally used to decrypt and access the EMR data.

V. IMPLEMENTATION

This section demonstrates the novel blockchain-based architecture to securely share and access EMR with quantum-resistant cryptographic algorithms.

A. Interaction of System Components

The proposed architecture of the system is illustrated in Fig. 4 for blockchain-based EMR sharing using post-quantum cryptography. The sequence of actions is as follows:

1. User registers in the system, generates a cryptographic key pair – private and public key, and sends signed credentials to the Certificate Authority (CA).
2. CA issues a certificate on successful verification.
3. A Healthcare Professional (HP) creates an EMR and requests patient consent to upload patient data on the blockchain.
4. The patient may grant complete or partial consent or deny it. HP retrieves the patient's certificate and access policy if consent is granted.
5. HP encrypts the EMR data and uploads it on the IPFS network.
6. A reference to the uploaded EMR data is returned to the HP as Reference ID (RID).
7. HP signs a hash of the encrypted data, creates a transaction, and submits it to the blockchain.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

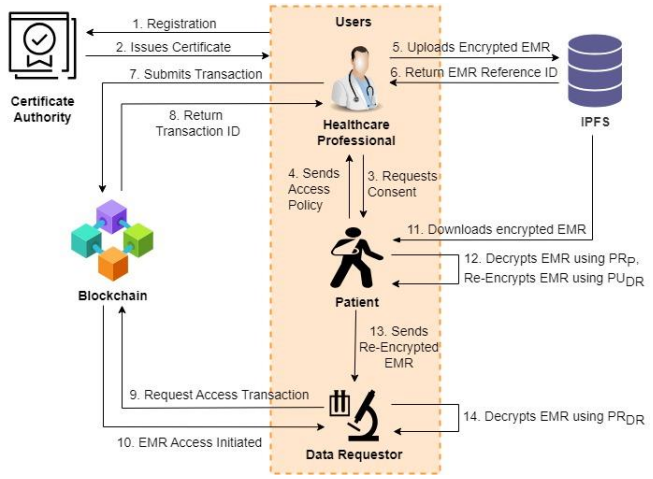


Fig. 4 Proposed Architecture

8. Once the transaction is validated after consensus, it is added to the blockchain and returns a transaction ID (TxID).
9. Data Requestor (DR) requests access to an EMR providing their certificate and TxID of the required EMR. The access request is verified based on the patient's access policy stored on the blockchain.
10. If the DR is authorized, the patient and the data requestor are notified.
11. Patient downloads the encrypted EMR data from IPFS using the RID stored in blockchain transaction.
12. Patient calculates the hash of the downloaded data and compares it with the signed hash stored in the transaction. If data integrity is verified, patient decrypts the EMR using his private key, generates a temporary key pair using DR's public key and then re-encrypts the decrypted EMR using the temporary symmetric key.
13. Patient sends the re-encrypted EMR data and the temporary cipher key to the DR.
14. DR uses their private key to decapsulate the temporary cipher key, which is used to decrypt the re-encrypted EMR data and gains access to the patient's medical record.

B. User Registration

Every user has to first register in the system. The user generates a quantum-resistant key pair composed of a public key (PU) and a private key (PR) using the Kyber algorithm. It is only executed once for every new user. Then, the user combines the public key (PU) and an identification document as his credentials and signs it with his private key. The user credentials and the signature are sent to the CA for verification. The CA issues a certificate encapsulating the user's public key and User ID (ID) after successful verification. The certificate is signed by the private key of the CA and sent to the user. If the verification fails, the CA rejects the registration request. Algorithm 1 illustrates the pseudocode for user registration.

C. Record Creation

HP generates a patient's EMR and requests the patient's permission to upload the EMR data to the blockchain. The patient reviews the request and may grant complete or partial consent or deny the permission to upload the EMR data. If consent is granted, the patient will provide his access policy and certificate containing his public key. HP uses a hybrid encryption approach to encrypt the EMR. First, a random symmetric key is generated and encapsulated with the patient's public key into a cipher key (encrypted symmetric key) using the Kyber algorithm. Then, the actual EMR is encrypted with the symmetric key using the AES-256. The encrypted EMR data is uploaded to the IPFS network, and the IPFS returns a reference ID (RID) of the encrypted EMR data. The hash of the encrypted data is calculated and HP signs it with his private key. A transaction is created that contains the signed hash, RID of the uploaded EMR on IPFS, cipher key, patient's access policy, and other relevant metadata (timestamp, data type, etc.) and submitted to the blockchain network. If the transaction details are validated and consensus is reached, it is recorded in the blockchain. This process is described in Algorithm 2.

Algorithm 1: Registration of a user Pseudocode (registerUser)
{PU, PR} ← generate_keypair()
credentials = {PU, ID}
S1 ← signData(PR, credentials)
Send {S1, credentials} to CA
IF verify(credentials, S1, PU) == **TRUE**
  Verification is successful
  certificate ← issueCertificate(PU , ID)
  S2 ← signData(PR_{CA}, certificate)
  RETURN {S2, certificate}
ELSE
  Display an error message

Algorithm 2: Creation of an EMR transaction on blockchain pseudocode (createEMR)
EMR ← generateRecord()
IF consentGranted(patientCertificate.ID) == **TRUE**
  patientPolicy ← getPatientPolicy(patientCertificate.ID)
  SK, EK ← kyberEncrypt(patientCertificate.PU)
  encryptedEMR ← AESEncrypt(EMR, SK)
  RID ← uploadToIPFS(encryptedEMR)
  h ← hash(encryptedEMR)
  s_h ← signData(PR_{HP}, h)
  transData ← createTransaction(s_h, h, RID, EK, patientPolicy)
  TxID ← submitTransaction(transData)
  RETURN TxID
ELSE
  Display error message

D. Access and Share Record

A Data Requestor (DR), such as a lab technician, medical researcher, or insurance company, identifies the data of interest and submits a request to access a patient's EMR. The system verifies the access request as per the patient's access policy stored in the blockchain transaction. If the data requestor is authorized, the patient and data requestor are notified to initiate the process. The patient retrieves the RID from the blockchain transaction. RID is used to locate and access the encrypted EMR data from the IPFS network. Then, the patient downloads the encrypted EMR from IPFS using the RID. The hash H(EMR) of the downloaded EMR data is calculated. The patient retrieves the signed hash H'(EMR) from the blockchain transaction and compares it with the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

computed hash. Data integrity is verified if H(EMR) matches H'(EMR). Now, patient retrieves encrypted symmetric key from the blockchain transaction. Kyber decryption is performed using the patient's private key to decapsulate encrypted symmetric key. The encrypted EMR is decrypted using the symmetric key to obtain the EMR data.

To share the EMR data with the data requestor, patient generates a temporary symmetric key (SK_t) and corresponding cipher key (EK_t) pair with the public key of data requestor using Kyber encrypt algorithm. The decrypted EMR data is re-encrypted with SK_t. Re-encrypted EMR data and EK_t are transmitted to the data requestor. The data requestor uses its private key to decapsulate EK_t into SK_t, and SK_t is used to decrypt the re-encrypted EMR to access the desired medical record. Algorithms 3 and 4 describe the process for requesting access to an EMR and secure sharing of the desired EMR respectively.

Algorithm 3: Request for access to EMR data Pseudocode (RequestAccessEMR)

```
TransactionData ← getTransactionData(TxID)
IF isAuthorised(DRCertificate.ID,
TransactionData.patientPolicy) == TRUE
   NotifyPatient(patientCertificate.ID, "Access granted")
   ReEncryptedEMR, EKt ← getRequestedEMR(TxID,
DRCertificate)
   SKt ← KyberDecrypt(PRDR, EKt)
   DecryptedEMR ← AESDecrypt(ReEncryptedEMR, SKt)
   RETURN DecryptedEMR
ELSE
   NotifyPatient(patientCertificate.ID, "Access denied to Data
Requestor")
```

Algorithm 4: Grant access to EMR data for Authorized requestor Pseudocode (getRequestedEMR)

```
TransactionData ← GetTransactionData(TxID)
RID = TransactionData.RID
DownloadedEMR ← DownloadFromIPFS(RID)
DownloadedHash ← hash(DownloadedEMR)
Hash = TransactionData.Hash
IF DownloadedHash == Hash, THEN
   EK = TransactionData.CipherKey
   SK ← KyberDecrypt(PRP, EK)
   DecryptedEMR ← AESDecrypt(DownloadedEMR, SK)
      SKt, EKt ← KyberEncrypt(DRCertificate.PU)
      ReEncryptedEMR ← AESEncrypt(DecryptedEMR, SKt)
      RETURN ReEncryptedEMR, EKt
ELSE
      Display data integrity error message
```

VI. DISCUSSION

This section will explore the proposed architecture's resistance to quantum attacks, data security and privacy, data integrity, and patients' ownership of data. Table 1 provides a comparison of the features of the proposed model with the traditional centralized approaches.

A. Quantum Attack Resistance

The proposed post-quantum blockchain is a secure technology that utilizes blockchain technology and quantum-resistant cryptography. It protects the architecture from classical attacks and ensures protection against known quantum attacks, such as Shor's and Grover's algorithms, ultimately improving security. The proposed architecture uses a lattice-based PQC algorithm for key encapsulation, CRYSTALS-Kyber. Lattice-based cryptosystems embrace lattices to address NP-hard problems which are confirmed to be safe for both classical and quantum computers [15]. The hybrid encryption technique of CRYSTALS-Kyber and AES-256 is rendered as a practical approach for medical data sharing, which offers secure protection.

B. Data Security and Privacy

The use of blockchain technology enhances data immutability, transparency, and tamper-proof auditing. With CRYSTALS-Kyber integrated, EMR data is secured against potential attacks from quantum computers and enhances the underlying cryptographic security levels relative to conventional methods. By storing the EMRs off-chain, on a decentralized storage network, the risks of a single point of access failure and unauthorized access is eliminated. The data retrieval-re-encryption process with temporary keys offers added layers of security and privacy during access from authorized requestors. Lastly, the system maintains data integrity by storing hashes of the encrypted EMR data on the blockchain, allowing patients to verify its authenticity before decrypting.

C. Data Ownership and Access Control

The architecture is designed for patients to explicitly grant consent before any EMR data is uploaded to the blockchain. They have the right to control the extent of data sharing. The system emphasizes patient control by enabling them to define their access policies for specific data elements within their EMRs to specify who can access their data. The paper proposes a patient-centric model where the patient's complete ownership of the medical data resides.

Table 1 Comparison of the proposed model with traditional architectures

Aspect	Centralized Traditional Architectures	Proposed Quantum-resistant Blockchain-based Architecture
Security	Vulnerable to quantum attacks (pre-quantum cryptography used)	Quantum-resistant cryptography (CRYSTALS-Kyber) is used for improved security.
Data Immutability	Data can be potentially altered or deleted.	Blockchain ensures tamper-proof and immutable data storage.
Data Privacy	Privacy breaches are possible due to centralized control.	Off-chain data storage with access control and re-encryption are used for enhanced privacy.
Data Ownership	Patients may have limited or no control over their data.	Patient-centric model with explicit consent and access policy management is proposed for data sharing.
Scalability	Struggles with large data volumes	Potential for scalability through hybrid data storage and permissioned blockchain

VII. CONCLUSION

With the dynamic development of quantum computing, sensitive data cannot be trusted with pre-quantum cryptography. The paper seeks to address the threat of a potential attack from a quantum computer by proposing a novel blockchain architecture integrated with quantum-resistant cryptography for secure EMR sharing. The tamper-proofing and transparency of blockchain technology assures data immutability. The hybrid encryption mechanism using

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

CRYSTALS-Kyber and AES-256 strengthens security against cryptographic security threats. Furthermore, the incorporation of PQC encapsulated within the hybrid encryption mechanism offers to protect sensitive EMR data against potential quantum attacks. Patient control of all data is also given priority to comply with data privacy. Patients will own their EMR completely and be able to determine the people they want to grant access to their data through access policy management. Authorized data requesters are granted access to desired EMRs through a secure re-encryption mechanism. Thus, the proposed architecture effectively deals with all the issues of ensuring data security, patient privacy, ownership control, and secure data sharing in the modern healthcare ecosystem.

However, quantum computing is rapidly developing, and even today's quantum-resistant encryption may become unsafe in the future. Therefore, it is necessary to develop rigorous security measures for both on-chain and off-chain data storage. There is a need to explore the regulatory and ethical implications of implementing quantum-resistant blockchain for EMR data sharing. Future research may include work on integrating the proposed architecture into existing healthcare systems and evaluating its performance by comparing it against existing solutions regarding scalability, efficiency and security. Altogether, the proposed architecture makes a significant contribution to EMR sharing securely and lays the foundation for a more secure healthcare data management system in the quantum computing era.

REFERENCES

[1] I. Keshta and A. Odeh, "Security and privacy of electronic health records: Concerns and challenges," *Egyptian Informatics Journal*, vol. 22, no. 2, pp. 177–183, Jul. 2021, doi: 10.1016/J.EIJ.2020.07.003.

[2] R. Venkatesh and B. S. Hanumantha, "A Privacy-Preserving Quantum Blockchain Technique for Electronic Medical Records," *IEEE Engineering Management Review*, 2023, doi: 10.1109/EMR.2023.3319376.

[3] H. Jin, Y. Luo, P. Li, and J. Mathew, "A Review of Secure and Privacy-Preserving Medical Data Sharing," *IEEE Access*, vol. 7, pp. 61656–61669, 2019, doi: 10.1109/ACCESS.2019.2916503.

[4] R. Venkatesh and B. Savadatti Hanumantha, "Electronic medical records protection framework based on quantum blockchain for multiple hospitals," *Multimed Tools Appl*, pp. 1–14, Oct. 2023, doi: 10.1007/S11042-023-16848-Y/FIGURES/D.

[5] "What is Quantum Computing? | IBM." Accessed: Mar. 27, 2024. [Online]. Available: https://www.ibm.com/topics/quantum-computing

[6] P. W. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer," *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1484–1509, Aug. 1995, doi: 10.1137/S0097539795293172.

[7] L. K. Grover, "A fast quantum mechanical algorithm for database search," *Proceedings of the Annual ACM Symposium on Theory of Computing*, vol. Part F129452, pp. 212–219, May 1996, doi: 10.1145/237814.237866.

[8] T. M. Fernandez-Carames and P. Fraga-Lamas, "Towards Post-Quantum Blockchain: A Review on Blockchain Cryptography Resistant to Quantum Computing Attacks," *IEEE Access*, vol. 8, pp. 21091–21116, 2020, doi: 10.1109/ACCESS.2020.2968985.

[9] M. Bhavin, S. Tanwar, N. Sharma, S. Tyagi, and N. Kumar, "Blockchain and quantum blind signature-based hybrid scheme for healthcare 5.0 applications," *Journal of Information Security and Applications*, vol. 56, p. 102673, Feb. 2021, doi: 10.1016/J.JISA.2020.102673.

[10] K. Kaushik and A. Kumar, "Demystifying quantum blockchain for healthcare," *Security and Privacy*, vol. 6, no. 3, p. e284, May 2023, doi: 10.1002/SPY2.284.

[11] M. K. Shrivas, S. Kachhwaha, A. Bhansali, and S. Vir Singh, "Quantum-resistant University Credentials Verification System on Blockchain," *Proceedings of the 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development, NIGERCON 2022*, 2022, doi: 10.1109/NIGERCON54645.2022.9803153.

[12] S. Gao, D. Zheng, R. Guo, C. Jing, and C. Hu, "An anti-quantum e-voting protocol in blockchain with audit function," *IEEE Access*, vol. 7, pp. 115304–115316, 2019, doi: 10.1109/ACCESS.2019.2935895.

[13] P. Zhang, L. Wang, W. Wang, K. Fu, and J. Wang, "A Blockchain System Based on Quantum-Resistant Digital Signature," *Security and Communication Networks*, vol. 2021, 2021, doi: 10.1155/2021/6671648.

[14] D. Gurung, S. R. Pokhrel, and G. Li, "Performance Analysis and Evaluation of Post Quantum Secure Blockchained Federated Learning," Jun. 2023, Accessed: Mar. 28, 2024. [Online]. Available: https://arxiv.org/abs/2306.14772v2

[15] S. Gupta, K. K. Gupta, P. K. Shukla, and M. K. Shrivas, "Blockchain-based Voting System Powered by Post-Quantum Cryptography (BBVSP-PQC)," in *ICPC2T 2022 - 2nd International Conference on Power, Control and Computing Technologies, Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2022. doi: 10.1109/ICPC2T53885.2022.9776966.

[16] D. Dhinakaran, D. Selvaraj, N. Dharini, S. E. Raja, and C. S. L. Priya, "Towards a Novel Privacy-Preserving Distributed Multiparty Data Outsourcing Scheme for Cloud Computing with Quantum Key Distribution," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 2, pp. 286–300, Dec. 2023, Accessed: Mar. 28, 2024. [Online]. Available: https://www.ijisae.org/index.php/IJISAE/article/view/4252

[17] "What Is Blockchain? | IBM." Accessed: Mar. 28, 2024. [Online]. Available: https://www.ibm.com/topics/blockchain

[18] I. Bashir, "Mastering Blockchain," *Packt*, vol. 161, no. 4, pp. 335–339, 2017, Accessed: Mar. 28, 2024. [Online]. Available: https://www.packtpub.com/product/mastering-blockchain/9781787125445

[19] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System." [Online]. Available: www.bitcoin.org

[20] A. Karakaya and A. Ulu, "A Review on Latest Developments in Post-Quantum Based Secure Blockchain Systems," *ISDFS 2023 - 11th International Symposium on Digital Forensics and Security*, 2023, doi: 10.1109/ISDFS58141.2023.10131840.

[21] "Designing a Blockchain Architecture: Types, Use Cases, and Challenges | by Svetlana Cherednichenko | mobindustry | Medium." Accessed: Mar. 28, 2024. [Online]. Available: https://medium.com/mobindustry/designing-a-blockchain-architecture-types-use-cases-and-challenges-9894fb7b58e

[22] N. Szabo, "Formalizing and Securing Relationships on Public Networks," *First Monday*, vol. 2, no. 9, Sep. 1997, doi: 10.5210/FM.V2I9.548.

[23] Y. L. Gao, X. B. Chen, Y. L. Chen, Y. Sun, X. X. Niu, and Y. X. Yang, "A Secure Cryptocurrency Scheme Based on

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

Post-Quantum Blockchain," *IEEE Access*, vol. 6, pp. 27205–27213, Apr. 2018, doi: 10.1109/ACCESS.2018.2827203.

[24] D. Micciancio and O. Regev, "Lattice-based Cryptography," *Post-Quantum Cryptography*, pp. 147–191, Jan. 2009, doi: 10.1007/978-3-540-88702-7_5.

[25] J. Dey and R. Dutta, "Progress in Multivariate Cryptography: Systematic Review, Challenges, and Research Directions," *ACM Comput Surv*, vol. 55, no. 12, Mar. 2023, doi: 10.1145/3571071.

[26] H. Singh, "Code based Cryptography: Classic McEliece," 2020.

[27] "What is Hash-based Cryptography? - Utimaco." Accessed: Mar. 28, 2024. [Online]. Available: https://utimaco.com/service/knowledge-base/post-quantum-cryptography/what-hash-based-cryptography

[28] G. Alagic *et al.*, "Status Report on the First Round of the NIST Post-Quantum Cryptography Standardization Process," Jan. 2019, doi: 10.6028/NIST.IR.8240.

[29] "Post-Quantum Cryptography | CSRC." Accessed: Mar. 28, 2024. [Online]. Available: https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022

[30] J. Bos *et al.*, "CRYSTALS - Kyber: A CCA-Secure Module-Lattice-Based KEM," *Proceedings - 3rd IEEE European Symposium on Security and Privacy, EURO S and P 2018*, pp. 353–367, Jul. 2018, doi: 10.1109/EUROSP.2018.00032.

[31] L. Ducas *et al.*, "Crystals-dilithium: A lattice-based digital signature scheme," *IACR Transactions on Cryptographic Hardware and Embedded Systems*, vol. 2018, no. 1, pp. 238–268, 2018, doi: 10.13154/TCHES.V2018.I1.238-268.

[32] "Falcon." Accessed: Mar. 28, 2024. [Online]. Available: https://falcon-sign.info/

[33] D. J. Bernstein, R. Niederhagen, A. Hülsing, J. Rijneveld, S. Kölbl, and P. Schwabe, "The SPHINCS+ Signature Framework," *Cryptology ePrint Archive*, pp. 2129–2146, Nov. 2019, doi: 10.1145/3319535.3363229.

[34] "Kyber." Accessed: Mar. 28, 2024. [Online]. Available: https://pq-crystals.org/kyber/

[35] P. Zeng, S. Chen, and K. K. R. Choo, "An IND-CCA2 secure post-quantum encryption scheme and a secure cloud storage use case," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, pp. 1–15, Dec. 2019, doi: 10.1186/S13673-019-0193-6/FIGURES/6.

[36] A. Aikata, A. Can Mert, M. Imran, S. Pagliarini, and S. Sinha Roy, "KaLi: A Crystal for Post-Quantum Security using Kyber and Dilithium".

[37] W. Wang, Y. Yu, and L. Du, "Quantum blockchain based on asymmetric quantum encryption and a stake vote consensus algorithm," *Scientific Reports 2022 12:1*, vol. 12, no. 1, pp. 1–12, May 2022, doi: 10.1038/s41598-022-12412-0.

[38] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," Jul. 2014, Accessed: Mar. 26, 2024. [Online]. Available: https://arxiv.org/abs/1407.3561v1

[39] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, "Blockchain for Secure EHRs Sharing of Mobile Cloud Based E-Health Systems," *IEEE Access*, vol. 7, pp. 66792–66806, 2019, doi: 10.1109/ACCESS.2019.2917555.

[40] D. L. Evans and K. H. Brown, "FIPS 197 Federal Information Processing Standards Publication Advanced Encryption Standard (AES)", doi: 10.6028/NIST.FIPS.197-upd1.

[41] M. N. Alenezi, H. K. Alabdulrazzaq, and N. Q. Mohammad, "Symmetric Encryption Algorithms: Review and Evaluation Study," *International Journal of Communication Networks and Information Security (IJCNIS)*, vol. 12, no. 2, 2020, doi: 10.17762/IJCNIS.V12I2.4698.

[42] M. Anastasova, P. Kampanakis, and J. Massimo, "PQ-HPKE: Post-Quantum Hybrid Public Key Encryption," *Cryptology ePrint Archive*, 2022.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 04:35:39 UTC from IEEE Xplore. Restrictions apply.

Check for updates

Received: 25 March 2024 | Revised: 12 June 2024 | Accepted: 20 June 2024

DOI: 10.1002/crq.21444

RESEARCH ARTICLE

Association for Conflict Resolution WILEY

Quitting silently: A longitudinal research on the impact of workplace conflict and nonviolent work behavior

Suhans Bansal | **Naval Garg**

Department of University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India

Correspondence
Naval Garg, Department of University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India.
Email: naval.garg@dtu.ac.in

Abstract

This study aimed to explore the moderation effect of nonviolent work behavior on the relationship between workplace conflict and quiet quitting. The study is based on a longitudinal research design and uses AMOS to explore the results. Age and gender were used as control variables. The results suggest that the two dimensions of workplace conflict, that is, relationship and task conflicts, lead to the development of quiet quitting among Indian IT industry employees. The results also indicate that practicing nonviolent work behavior can reduce the development of quiet quitting arising out of workplace conflict. To the best of the authors' knowledge, this is one of the first studies that confirm the moderating role of nonviolent work behavior on the relationship between workplace conflict and quiet quitting, especially in India.

1 | INTRODUCTION

The workplace conflict arise among employees due to differences in perspectives, personalities, aspirations, cultural backgrounds, values, and communication styles (Cvenkel, 2020) and in some cases due to organizational and situational factors (Leiß & Rausch, 2022). Researcher argued "an inverted U shaped" graph of conflict and performance, wherein low level of conflict results in positive outcomes and conflict beyond a certain level produces undesirable outcomes (Pereira et al., 2023). The favorable results of workplace conflict include work engagement (Ter Hoeven & Van Zoonen, 2020), job satisfaction, better decision-making, and higher performance (Chakraborty et al., 2022; Pradhan & Jena, 2016), happier employees (Chen & Francesco, 2003), and better work-life balance (Bhende et al., 2020). Whereas, negative consequences include

 © 2024 Wiley Periodicals LLC.

lower career commitment (Deshpande & Chakraborty, 2023), mental health issues (Bansal et al., 2023), poor team collaboration (Donohue et al., 2023), turnover intentions (Badar et al., 2023), promote workplace deviant behavior (Atatsi et al., 2019), degraded work performance (J, 2014). One such deviant behavior is quiet quitting. Arnetz et al. (2018) suggest that some employees prefer responding to any conflict or violence with violence, resulting in a "tit-for-tat" situation, on the other hand, some employees may respond with a milder yet powerful withdrawal behavior known as quiet quitting (Ahmed et al., 2013). Quite quitting does not mean quietly quitting the present assignment and look for alternative jobs, it refers to significant reduction in the work efforts and limit oneself to only explicitly defined job profile. It is "sometimes conceptualized not as a way for employees to leave their current positions but to redistribute and re-prioritize how their time is spent" (Anand et al., 2023). It is a psychological withdrawal that generally precedes turnover (Hamouche et al., 2023). Previous studies have linked several factors as causes of quiet quitting, including COVID-19 (Hamouche et al., 2023), job burnout (Chakraborty et al., 2022; Cox et al., 2017), throat-cut competition (Xiong & Chen, 2018), and imbalanced work-life (Tapper, 2022). Further, quiet quitting results in various detrimental consequences ranging from employees becoming isolated to higher employee attrition and reduced overall productivity for the firm (Li et al., 2021). Current studies have shown that employees who tend to indulge in quiet quitting have faced actions such as being labeled as underperformers, demotions, and, in severe cases, termination from job (Serenko, 2023). Nevertheless, this concept is still new to researchers and requires further exploration, especially in the context of workplace conflict.

1.1 | Importance of studying quiet quitting

The research on quiet quitting is limited, especially on its relationship with workplace conflict. It is important to examine these initial and mild forms of employees response to conflict for three important reasons. First, early stage behaviors like quite quitting offer clear clues about latent workplace conflict which, if not addressed immediately, may blow up into full scale conflict (Serenko, 2023). Thus, organizations can take clues from these early signs to avoid future troubles. Secondly, quite quitting largely gets unnoticed as employees perform his/her assigned or explicitly defined duties. But they seriously dent individual and organizational performance as it kills organizational creativity and innovation, knowledge sharing, employee engagement, and committment (Anand et al., 2023; Lu et al., 2017). Last, although such behaviors evades organizational scrutiny, the gossip they generate motivates others to follow the suit causing serious ramifications for any organization. Therefore, this study aims to examine the influence of workplace conflict on quite quitting.

1.2 | Intervention suggested

The role of individual level intervention to counter the occurrences of the employee deviant behaviors needs to be explored (Biswas & Chakraborty, 2023). One such intervention is nonviolent work behaviors. Nonviolent work behavior does not mean mere absence of violence, rather, it includes trust, empathy, mutual respect, and forgiveness in mutual interactions, ensuring that even violent actions are responded in positive and constructive manner, defusing the conflicting situation with peace, love, and trust. Nonviolent work behavior interventions promote a strong positive organizational culture and are both preventive and cost-effective (Lata & Garg, 2023).

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

According to Sarkar and Palo (2019), nonviolence is "more than nonappearance of violence or non-harming. Rather, it is a condition of mental ablution and positive acts through body, mind, and speech." Bhalerao and Kumar (2015) offer a definition of nonviolence in the workplace as an individual's propensity to respond constructively to any provocation, which might include a violation of respect norms and justice in the organization. The authors further asserted that it includes nonviolence (not harming others) in one's thoughts, words, and deeds (Bhalerao & Kumar, 2015; Biswas & Chakraborty, 2023). However, being a new concept, not enough studies have adequately explored the affects of nonviolent work attitude on employee turnover, degraded work motivation, and quiet quitting. Hence, the present study examines the moderating role of nonviolent work behavior on the relationship between workplace conflict and quite quitting.

1.3 | Theory used

Based on Bandura's (1986) social cognitive theory (SCT), it is proposed that workplace conflict (environmental factors) leads to quiet quitting (behavioral factors). According to this theory, behavioral, personal, and environmental factors are in a loop, with each factor affecting others (Bandura, 1986). However, their nonviolent work attitudes (personal factors) can affect their choice of response. Hence, they may see that conflict may have occurred due to various factor like job stress, inability to cope with job related demand among other factor (Bhalerao & Kumar, 2015), thereby, people may not resort to quiet quitting. With this the study seeks to answer the following questions:

Does workplace conflict promotes quiet quitting among employees?
Does practicing nonviolent work behavior during workplace conflict reduce quiet quitting?

1.4 | Contributions

In light of the discussion presented in the preceding paragraphs, the current study has three important contributions. First, it explores one of the recent workplace phenomena of quite quitting. Second, it examines quite quitting in the larger organizational context, that is, workplace conflict. Third, this study suggests that nonviolent work behavior moderates the relationship between workplace conflict and quite quitting, which to the best of our knowledge, none of the previous studies have evaluated.

This paper is divided into three sections. The first section elaborates the theoretical framework of the study. The second section explains the methodology adopted to test the study's hypotheses. And lastly, results and implications along with limitations and scope for future studies are elaborated.

1.5 | Theoretical framework

1.5.1 | Workplace conflict and quiet quitting

Researchers suggest conflict is of two types: relationship and task. Relationship conflict is between people, and task conflict is any conflict related to a task being performed/planned

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

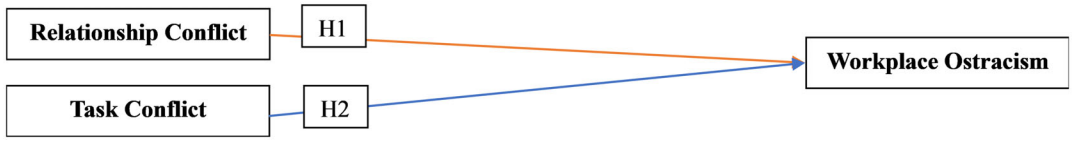


FIGURE 1 Proposed Relationship. *Source*: Bansal, S., Garg, N.

(Bhalerao & Kumar, 2015). Relationship conflict, one of the most common and widely occurring conflicts, can be triggered by various reasons, including poor communication, differences in belief systems, cultural differences, and the spreading of negative gossip in the workplace (Khan et al., 2021). Further, relationship conflict leads to an increase in the perceived job stress (Tu et al., 2022), for example, the team's work may require conflicting employees to work together, thus increasing their perceived job stress. This may further exacerbate workplace conflict and the development of employee deviant behaviors such as quiet quitting. Such behaviors can act as the early symptoms of latent conflict or severe consequences such as employee attrition and hence are required to be examined closely. It follows from the SCT that relationship conflict acts as one of the environmental factors that affect an individual's behavior, as a result of which they may tend to resort to quiet quitting (behavioral factor). Accordingly, following hypothesis and theoretical model (refer to Figure 1) are proposed.

H1. Relationship conflict positively relates to quiet quitting.

Similarly, task conflict is also one of the widely occurring conflicts in the workplace. According to Liao et al. (2022), task conflict can occur when employees are unclear about their responsibilities and authorities, and they are not sure about what is to be done and/or who will perform a specific task. Further, task conflict can be caused by the lack of proper communication, sudden change of goals, ambiguous information, change in requirements, or incomplete information (Shin et al., 2020). In such situations, if the concerned authority does not provide correct information, employees may perform only those tasks mentioned in their job description, thereby reducing their productivity. Serenko (2023) associates this situation with quiet quitting. Therefore, as per the SCT, individual's behavioral factors (quiet quitting) will be affected by their environmental factors (task conflict). Therefore, based on this theoretical premise, we propose the following hypothesis and theoretical model (see Figure 1).

H2. Task conflict positively relates to quiet quitting.

Although relationship conflict affects more severely than task conflict, researchers suggest that their combined effect is more detrimental to organizations and employees both. Hence, the need of the hour is to develop interventions to reduce their combined detrimental effect (Caesens et al., 2019). In other words, conflict and quiet quitting result in a phenomenon known as social death, that is, loss of social connectedness. Therefore the following moderation effect is proposed as an intervention.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

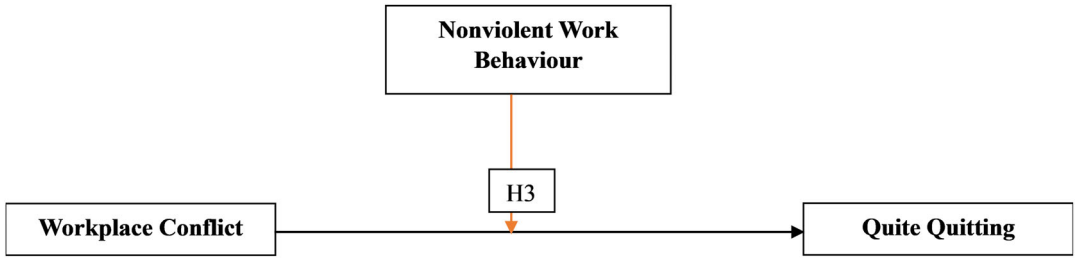


FIGURE 2 Proposed moderation effect. *Source*: Bansal, S., Garg, N.

1.5.2 | The moderating effect of nonviolent work attitudes

Bhalerao and Kumar (2015) proposed four dimensions of nonviolent work behavior. NVT-Ignore behavior, employees consider acts of violence as non-worthy and respond by ignoring them. NVT-Stress, employees give the benefit of the doubt to the aggressor and do not respond violently considering the acts of violence as a result of stress on the aggressor. NVT-ability, any violence is labeled as incompetence and the inability of the aggressor against the given task, therefore, the victim does not respond. NVT-constructive, victim responds indirectly to the violence by bringing out changes in self, thereby responding pro-socially and defusing the spiral of violence. Hence, through these dimensions, it is clear that violence is a spiraling situation wherein the violence increases when the victim responds aggressively to the provocations.

The proposed moderating effect of nonviolent work attitude amidst the relationship between workplace conflict and quite quitting is theoretically premised in the SCT meaning that workplace experiences and personal attitude defines one's behavior. Bandura (1991) suggested that one's behavior is regulated by environmental circumstances and personal factors like attitude. Apart from individual effects, the interaction of these factors also influence one's behavior (Chakraborty & Biswas, 2021). Therefore, we propose that nonviolent work behaviors (personal factors) can help an individual understand the occurrence of the workplace conflict and prevent them from resorting to quiet quitting. Based on the theoretical premise, following conceptual model (refer to Figure 2) and hypothesis are proposed.

> **H3.** Nonviolent work attitude moderates the relationship between workplace conflict and quiet quitting.

2 | METHODOLOGY

Based on Sawyer et al. (2022) recommendations to conduct a longitudinal research design for better generalization of the results, this study explores the problem of quiet quitting arising from workplace conflict. It also explores the moderating effect of the nonviolent work behaviors on the proposed relationship.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.1 | Sample and data collection

This study uses the data was collected from the Indian IT professionals located in the Delhi NCR region. They were approached through their HR managers. The information technology industry (ITI) is a closely knit workplace, where employees from diverse backgrounds having diverse work opinions, ethics and believes are employed (Garg & Saini, 2023). Other than workforce diversity, factors such as continuous learning, collaboration, innovation and problem solving, and stressing work comprise the ITI work environment (D'Cruz et al., 2021). Researchers suggest that a fine balance exists among these factors, which maintains harmony in the work environment (Garg & Saini, 2023). However, considering the nature of the ITI's work environment, studying the effect of workplace conflict and quiet quitting is important, because they can cause a disbalance in the environment. For instance, Parayitam et al. (2020) highlights workplace conflict as one of the major factors leading to the development of the role ambiguity in the ITI employees. Also, quiet quitting hinders organizational culture which impacts the knowledge sharing and collaborative environment in the it is (Shah & Hashmi, 2019). Therefore, the Indian IT industry was considered as the best suited for this study.

A two-day seminar on work environment and toxicity was conducted. Every participant was given a unique ID. The data was collected using a structured questionnaire via Google forms was shared across groups created for the seminar. To qualify as the participant for our study, it was taken account that an employee pose at least 3 years of work experience and has worked at least on five projects to which they must have contributed significantly. It was found that off the 632 seminar attendees, only 593 qualified for our study. At time T1, demographic and workplace conflict information were collected. After 3 months (T2), information on nonviolent work behavior was captured. Finally, after another 3 months (T3), quiet quitting intentions were assessed. Only 519 participants out of 593 qualifying attendees participated in the first wave of the research, that is, T1. Subsequently, 428, and 395 working professionals responded at times T2 and T3, respectively. Cochran (1977) has recommended a sample size of 384 at a 0.05 significance level, thus the sample size of the present study seems appropriate. Table 1 reports the demographic information of the sample.

2.2 | Measures

2.2.1 | Workplace conflict

Meier et al.'s (2013) relationship and task conflict scale (RTCS) was used to measure perceived workplace conflict. The scale was initially developed by Jehn (1995) and was reworded by Meier et al. (2013) to have positive statements rather than questioning statements. This eight-item scale measures two dimensions of conflict- relationship conflict and task conflict—on a seven-point rating scale ranging from one (not at all) to seven (extremely) for relationship conflict and between one (never) and seven (always) for task conflict. A few items are "there is a friction among members of my organization" and "there are differences of opinions regarding work within my organization/institution."

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Demographic profile.

Variables		*n*
Age	Below 25 years	25
	25–30 years	198
	30–35 years	132
	Above 35 years	40
Gender	Male	200
	Female	195
Relationship status	Single	168
	Married	107
	Committed	101
	Widowed	19
Education qualification	Under Graduate	109
	Post Graduate	248
	PhD	38
Work experience	Below 5 years	98
	5–10 years	157
	10–15 years	86
	Above 15 years	54

Source: primary data.

2.2.2 | Nonviolent work attitude

The study used adopted version of nonviolence at workplace scale developed by Bhalerao and Kumar (2015). It is a 32-item scale that rates statements on a five-point rating scale ranging from one (strongly disagree) to five (strongly agree). A few items of the scale are "I inquire if the subordinate is upset due to some reason if he/she threatens me" and "I overlook impoliteness of co-workers."

2.2.3 | Quiet quitting scale

The scale has been taken from Anand et al.'s (2023) quiet quitting and quiet firing scales. The adopted scale is a unidimensional seven-item five-point rating scale ranging between one (completely disagree) and five (completely agree). The scale is used to measure the quiet quitting intentions of employees. A few items are "I am doing bare minimum work to avoid being fired" and "I feel I have a lack of meaningfulness at work."

2.3 | Reliability, validity, multicollinearity, and common method bias

The internal consistency reliability was assessed using Cronbach's alpha (CA), composite reliability (CR), and McDonald's omega (MO) values. The recommended values of these measures

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Descriptive statistics and discriminant validity.

Variables (study 2)	Mean	SD	CA	CR	MO	AVE	Discriminant validity RC	Discriminant validity TC
Relationship conflict	3.85	0.211	0.825	0.917	0.800	0.53	1 (0.53)	
Task conflict	4.45	0.153	0.839	0.942	0.817	0.55	0.362	1(0.55)
Workplace conflict	4.60	0.192	0.801	0.916	0.821	0.56		
Nonviolent work behavior	3.24	0.210	0.812	0.918	0.798	0.61		
Quiet quitting	4.80	0.351	0.807	0.911	0.811	0.59		

Source: primary data; RC–relationship conflict; TC–tack conflict.

are 0.70 and above (George & Mallery, 2003; Hayes & Coutts, 2020). Table 2 reports that CA, CR, and MO values for relationship conflict, task conflict, overall workplace conflict, nonviolent work behavior, and quite quitting are above 0.70, which established internal consistency reliability of the scales. The average value explained (AVE) values were used to ensure convergent validity of the scales. The accepted value of AVE is 0.50 and above. Table 2 states that the AVE values for relationship conflict, task conflict, overall workplace conflict, nonviolent work behavior, and quite quitting are 0.53, 0.55, 0.56, 0.61, and 0.59. The discriminant validity was assessed using the Fornell and Larcker (1981) testing system, which states that the squared correlation between the dimensions of the scales should be less than their AVE values. Again, Table 2 concluded the discriminant validity of the workplace conflict scale. Since there are two dimensions of workplace conflict, the multicollinearity was adjudged with the help of the variance inflation index (VIF) values, whose acceptable values are less than three. The VIF values for relationship conflict and task conflict are 1.24 and 2.67, which confirmed absence of significant multicollinearity. This study is based on the longitudinal research design, which is poised to negate common method bias. However, certain procedural and statistical methods are employed to minimize it. The methods like usage of pre-established scales, promise of anonymity and confidentiality, informed consent, and strict academic use of collected data reduce chances of common method bias (Podsakoff et al., 2003). Further, when all items of the study were subjected to factor analysis, more than one factors emerged. Also, single largest factor explains only 31.27% variance, which negates concluded absence of significant common method bias according to the Harman's one-factor method.

2.4 | Analysis of data

The study's explored the intricacies between workplace conflict, quiet quitting, and nonviolent work behavior using structural equation modeling (SEM) in IBM's AMOS 21. Two models- causal model and moderation model- were evaluated for model fitness.

3 | RESULTS

Table 3 reports the correlation coefficients of the variables used in the study. It reports significant positive correlations between the two dimensions of workplace conflict and quiet quitting

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Correlation matrix.

Variables	1	2	3	4	5
Relationship conflict	1				
Task conflict	0.523*	1			
Workplace conflict	0.560*	0.611*	1		
Nonviolent work behavior	−0.271*	−0.214*	−0.330*	1	
Quiet quitting	0.467*	0.415*	0.477	−0.381	1

*Sig. at 0.05.
Source: primary data.

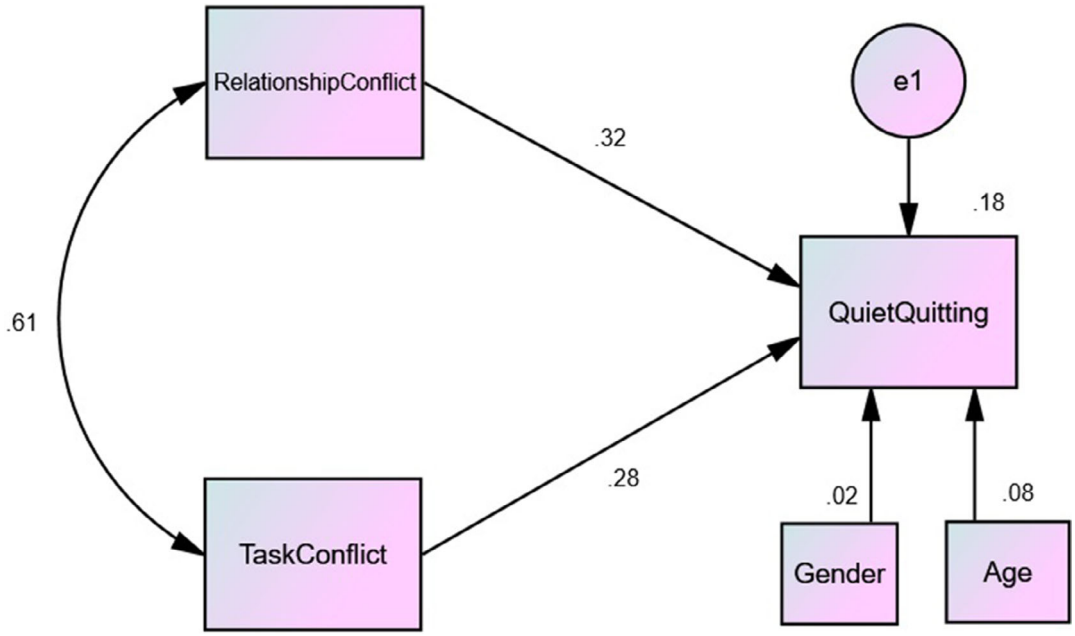


FIGURE 3 Model 1–proposed causal model. *Source*: Bansal, S., Garg, N.

($r_{\text{relationshipconflict}} = 0.467$, $p < 0.05$, and $r_{\text{taskconflict}} = 0.415$, $p < 0.05$). Thus, the results from correlation analysis suggest that workplace conflict positively affects quiet quitting, meaning workplace conflict can lead to higher quiet quitting intentions at the workplace. Further, the results of correlation analysis also suggest that nonviolent work attitude negatively correlates with the two dimensions of workplace conflict ($r_{\text{relationshipconflict}} = -0.271$, $p < 0.05$, and $r_{\text{taskconflict}} = -0.214$, $p < 0.05$) and quiet quitting ($r_{\text{quietquitting}} = -0.381$). These results are in line with previous studies, such as those conducted by Gürlek (2021), Sarkar and Garg (2020), and Guenter et al. (2016). Figure 1 represented a causal model that includes relationship conflict, task conflict, and quite quitting, which was evaluated using SEM in AMOS 21. As suggested by previous researchers, gender and age were controlled in the model. The threshold values of model fit indices are RMSEA <0.8, SRMR ≤0.05, GFI and TLI ≥0.95, and CFI ≥0.97 (Xia & Yang, 2018). Figure 3 represented SEM causal model. The model reports acceptable model fit values of RMSEA = 0.712, SRMR = 0.043, GFI = 0.972, TLI = 0.977, and

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

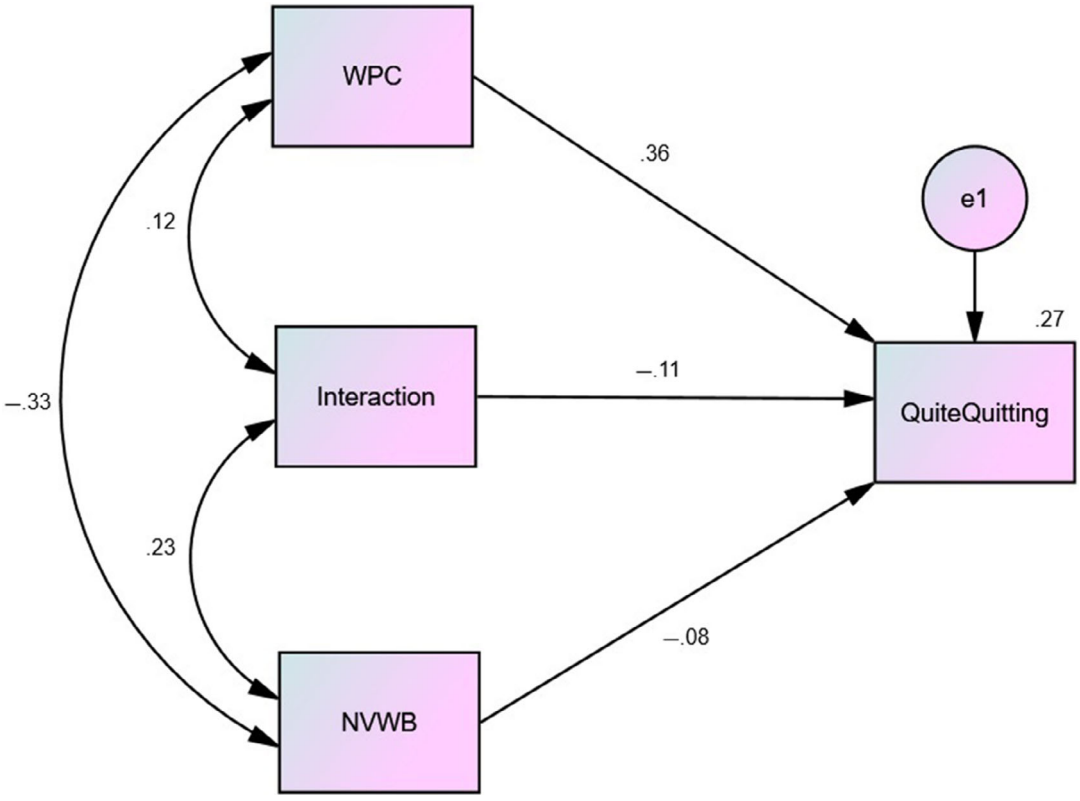


FIGURE 4 Model 2–Moderation analysis. *Source*: Bansal, S.; Garg, N.; WPC–workplace conflict, NVWB–nonviolent work behavior, interaction–WPC*NVWB.

CFI = 0.974. The model reported significant relationships between a) relationship conflict and quiet quitting ($\beta = 0.32$, $p < 0.05$, and b) task conflict and quiet quitting ($\beta = 0.28$, $p < 0.05$). Also, it stated insignificant effect of age and gender with regression weights of $\beta_{age} = 0.08$ and $\beta_{gender} = 0.02$. These results confirmed the acceptance of the first two hypotheses of the study. Figure 2 illustrates moderation model that examine the moderating effect of nonviolent work behavior amid the relationship between workplace conflict and quiet quitting. Figure 4 depicted moderation model in SEM, which reported a good fit overall with RMSEA = 0.561, SRMR = 0.036, GFI = 0.977, TLI = 0.986, and CFI = 0.989. It was interesting to note that the moderation model exhibited better model fit than the causal model, which suggest significant moderation effect of nonviolent work attitude. Workplace conflict and nonviolent work attitude were significantly related to quite quitting with βworkplaceconflict = 0.36 and βquitequitting = −0.08 at $p < 0.05$. The interaction effect of workplace conflict and nonviolent work attitude was also statistically significant with regression weight of −0.11 at $p < 0.05$. Hence, the moderation model is concluded to be statistically significant. Hence, H3 is also accepted. Figure 5 depicts diagrammatical representation of the moderation effect of nonviolent work behavior. It suggested that at higher levels of nonviolent work attitude the impact of workplace conflict on quite quitting is relatively less potent and effective.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

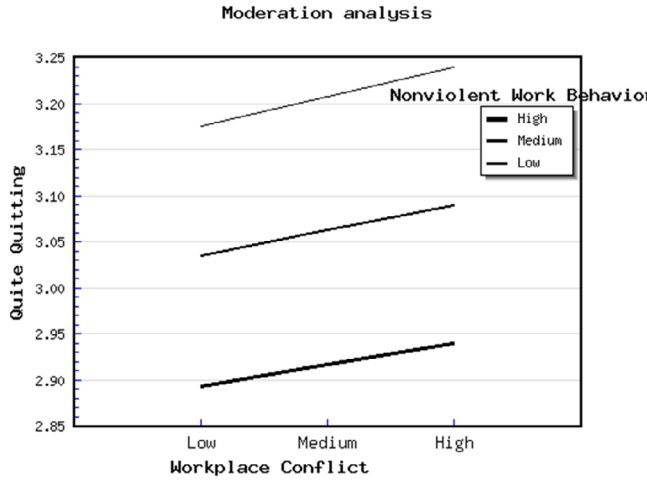


FIGURE 5 Moderation analysis. *Source*: primary data.

4 | DISCUSSION

4.1 | General discussion

The constant human-human interaction results in both positive or negative outcomes. Various social scientists and managers suggest that of these outcomes, conflict is one of the significant predictors of employee deviant behaviors (Sarkar & Garg, 2021). Researchers suggest that conflict is of two types: relationship and task (Bhalerao & Kumar, 2015), and can result in deviant behaviors ranging between mild to harsh deviant behaviors, such as lower productivity and absenteeism (Cong et al., 2015) to workplace bullying and harassment (Hamre et al., 2021). Though harsh deviant behaviors are of grave concern, studying mild deviant behaviors is a need of the hour because these mild deviant behaviors are the early warning signals of something of grave concern is happening or will occur in the near future. One such mild deviant behavior is quiet quitting (Serenko, 2023). As discussed in the earlier sections, quiet quitting can lead to the lower firm's productivity and can encourage higher employee absenteeism. Further, experts suggest when nonviolent behavior at the individual level are used for tackling such issues, they can benefits organizations and their employees because they are helpful in difusing the spiral of conflict without the need to intervene through punishments (Bhalerao & Kumar, 2015). Therefore, based on Bandura's (1986) SCT based interaction of personal, environmental, and behavioral factors, the study proposed that the two dimensions of workplace conflict significantly positively relate to quiet quitting intentions at work, and nonviolent work behavior can be used as a moderator with the proposed relationship. Moreover, based on the suggestions of various researchers, this study used a longitudinal research design. The data was collected in three waves. Further, the study used AMOS 21 to analyze the results of the data collected from the employees working in the Indian IT sector.

The proposed relationships between the two dimensions of workplace conflict, that is, relationship conflict, task conflict, and nonviolent work attitude, and quiet quitting were assessed using correlation analysis. The results suggested that relationship and task conflict have positive associations with quiet quitting, meaning, an employee can develop quiet quitting intentions if

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relationship and task conflict occur. The results further indicate that nonviolent work attitude negatively correlates with the two dimensions of workplace conflict and quiet quitting, indicating that nonviolent work attitudes can be used during conflict to avoid quiet quitting. However, the results of the correlation analysis and model fit of the proposed relationship and moderation effect were further tested with a more robust testing method using structural equation modeling in AMOS 21. Age and gender were used as the control variables in Model 1 and they reported insignificant results. This means that quiet quitting is a natural phenomenon and an individual can resort to it despite of their age and gender. The results of Model 1 also report that the two dimensions of workplace conflict are positively associated with quiet quitting. The results of this study align with the SCT and suggest that relationship and task conflicts (environmental factors) can lead to quiet quitting (behavioral factor) among employees as conflict-coping strategies. As a result, they may decrease their performance to the bare minimum required levels to safeguard their jobs, arrive late in meetings, or avoid people even during informal meeting times like lunchtime. However, they are neither helping themselves or their organizations nor resolving conflict. Hence, the spiral of conflict is not diffused. The results of our study are inline with previous studies like the ones conducted by Dukhaykh (2023) and Gaiji et al. (2021). Dukhaykh (2023) suggested that workplace conflict is one of the major causes of emotional drain and emotionally drained employees find it hard to focus on their official tasks. Also, the study conducted by Gaji et al. (2021) indicated that employees may psychologically withdraw from their work as a result of their coping mechanism against workplace conflict, and such employees may show withdrawal behaviors by working only the bare minimum required.

The study also assessed the use of nonviolent work attitudes to diffuse the spiral of workplace conflict and quiet quitting. For this reason, the current study explored the moderation effect of nonviolent work behavior on the relationship between workplace conflict and quiet quitting using SEM in AMOS 21. The results of the moderation analysis (Model 2) emphasize the concept of reciprocal determinism of the SCT, where nonviolent work attitudes (personal factors) interact with workplace conflict (environmental factors) and affect the outcome of the environmental factors. For instance, the value of the interaction effect ($\beta = -0.11$) indicates that the nonviolent work behavior is one of the suitable moderators that can be used to diffuse the spiral of conflict. This interaction affects the third factor, behaviourial. In other words, when people exhibit nonviolent work attitudes, they can effectively adjust their responses to workplace conflict, leading to reduced impact of workplace conflict on quiet quitting intentions. Moreover, the results suggest that individuals with higher nonviolent work attitudes can quickly adjust their beliefs and approach conflict resolution more constructively and confidently. Hence, it is safe to say that practicing nonviolent work behavior during workplace conflict can cushion quiet quitting intentions arising due to workplace conflict. Consequently, the results of this study highlight the need to promote and reinforce nonviolent work attitudes and strategies in the Indian IT sector. Moreover, the results of the moderation analysis are inline with earlier studies like the one conducted by Mahipalan and Garg (2024), where the authors have suggested that practicing nonviolence can reduce the ill effects of workplace toxicity on employees psychological capital. Hence, it will be helpful to practice nonviolence during workplace conflict, which can reduce the occurrence of quiet quitting, especially among IT employees.

4.2 | Implications

The study presents exciting findings on workplace conflict, quiet quitting, and nonviolent work behavior. It also has the following implications.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.2.1 | Theoretical implications

The present study extends the literature on workplace conflict and quiet quitting. It suggests a gap that indicates that no previous study has explored quiet quitting arising out of workplace conflict. Therefore, the findings of this study fill this gap. The results suggest that relationship and task conflicts are positively associated with quiet quitting, indicating that relationship and task-related conflicts in the workplace can lead to employee(s) resorting to quiet quitting. The results further indicate that workplace conflict can also be associated with an antecedent of quiet quitting. Second, the study also extends the literature on nonviolent work behavior. The results of this study empirically prove that nonviolent work behavior can act as a cushion against quiet quitting arising out of workplace conflict. Therefore, an employee may use nonviolence as a tool to diffuse the spiral of workplace conflict and quiet quitting as they may perceive conflict as an inability of other employee to cope with stress or job related demands, or may try to bring out positive change in the self. Third, this study further extends the literature on SCT. The results of our study empirically explain triadic reciprocal causation between the environmental, behavioral, and personal factors. The results indicate that workplace conflict (environmental factor) leads to the development of quiet quitting (behavioral factor) and when an employee practices nonviolent work behavior (personal factor), such behavior reduces the quiet quitting arising out of workplace conflict.

4.2.2 | Practical implications

First, the results of our study shed light on the importance of tackling the rise of workplace conflict. The results will help managers, academicians, and policymakers to develop policies, programs, and regulations to create awareness about workplace conflict and its detrimental effects on people. Recently, we conducted a seminar on the importance of the nonviolence in colleges. On similar grounds, managers may conduct such seminars that highlight the importance of practicing nonviolence at the workplace. Also, such seminars may inculcate the methods and practices that employees can develop to practice nonviolent behavior during workplace conflict. Second, the results will help organizations and institutions combat the quiet quitting issue. Various policies can be developed to identify cases of quiet quitting. Also, these policies may mandate conducting workshops and seminars on quiet quitting, its nature, how it may arise, and how it affects an individual, organization, and families. Third, organizations operating in Indian culture need to be aware of the quiet quitting tendency of employees and should work on creating policies that promote a balance between collectivism and individual well-being while preserving organizational cultural values. Fourth, the results will help managers, policymakers, and academicians to incorporate an open-door policy wherein people can approach their managers and discuss the conflict situation and how they feel about it. Even if an employees has been detected resorting to quiet quitting, the open door policy may allow the mangers to investigate the reasons for it as such behavior are the results of numerous factors like workplace conflict and burnout (Biswas & Chakraborty, 2023). Fifth, the scales used in this study can be utilized to assess workplace conflict and quiet quitting proactively, and managers can devise corrective measures to identify and counter these issues. Further, recruitment policies may be amended to adopt the scale measuring the nonviolent work behavior to assess the nonviolent behavior among the potential candidates. Such behavior can indicate that a nonviolence preferring candidate would always try to look the cause of the conflict and act upon it,

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

thereby, diffusing the spiral of conflict. Last, the results of our study promote the use of nonviolent work behavior to counter or reduce quiet quitting arising out of workplace conflict. The use of Ahimsa may be promoted by developing and deploying plans, policies, and regulations because our results empirically support the use of Ahimsa to intervene in workplace conflict and quiet quitting. Hence, policies based on nonviolent work behavior shall be developed. For instance, based on the critical incident appraisal method, a new employee appraisal policy can be developed where every employee would be appraised on their use of nonviolent work behaviors, especially during conflict situations.

4.3 | Limitations and future research

Though this study presents unique and exciting findings, it has certain limitations. First, the study is the first of its kind and uses a sample representing the Indian population. Though not a limitation in itself, future researchers are encouraged to study the proposed relationship and moderation effect in the cross-cultural environment to produce more insightful results. Second, future researchers may use other variables, such as gratitude, psychological capital, karmic law, etcetera, to moderate the proposed relationship. Moreover, future researchers may analyze the results of the proposed relationship against big five personality types or use other demographic variables as the control variables. Third, the study uses a combined conflict scale while exploring the moderation analysis of nonviolent work behaviors on the proposed relationship. Future researchers may explore the moderation effect of NVWB separately on the two dimensions of workplace conflict to produce more robust results. Fourth, the study has used workplace conflict as one of the predictors of quiet quitting. The researchers are encouraged to use variables like workplace bullying, job stress, and burnout, as predictors of quiet quitting. Last, as previous studies have shown that modern day organizations are multi-generational organizations, it is imperative to add how generational differences impact the use of nonviolence in the organization and quiet quitting intentions.

4.4 | Conclusion

In a nutshell, the study proposed and confirmed that workplace conflict leads to the development of quiet quitting among employees. The study also confirmed that nonviolent work behavior moderates the relationship between workplace conflict and quiet quitting, indicating that when a person practice/show nonviolent work behavior during workplace conflict, they defuse the spiral of conflict, thereby limiting the occurrence of quiet quitting.

AUTHOR CONTRIBUTIONS

Suhans Bansal–Written original draft, conceptualization, data collection, analysis. Dr. Naval Garg–Supervision, conceptualization, revising the original draft.

ACKNOWLEDGMENTS

We would like to thank all the participants of this study for taking out time from their busy schedules and become a part of our study. Also, we would like to appreciate their patience for being part of this study in all the time frames. Lastly, we would like to express our gratitude to reviewers of this manuscript as your valuable inputs have made this manuscript more impactful for the literature and research community.

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Suhans Bansal https://orcid.org/0000-0003-4682-0314
Naval Garg https://orcid.org/0000-0001-6694-9663

REFERENCES

Ahmed, I., Sultana, I., Paul, S. K., & Azeem, A. (2013). Employee performance evaluation: A fuzzy approach. *The International Journal of Productivity and Performance Management/International Journal of Productivity and Performance Management*, *62*(7), 718–734. https://doi.org/10.1108/ijppm-01-2013-0013

Anand, A., Doll, J. L., & Ray, P. (2023). Drowning in silence: A scale development and validation of quiet quitting and quiet firing. *The International Journal of Organizational Analysis*, *32*, 721–743. https://doi.org/10.1108/ijoa-01-2023-3600

Arnetz, J. E., Hamblin, L. E., Sudan, S., & Arnetz, B. B. (2018). Organizational determinants of workplace violence against hospital workers. *Journal of Occupational and Environmental Medicine*, *60*(8), 693–699. https://doi.org/10.1097/jom.0000000000001345

Atatsi, E. A., Stoffers, J., & Kil, A. (2019). Factors affecting employee performance: A systematic literature review. *Journal of Advances in Management Research*, *16*(3), 329–351. https://doi.org/10.1108/jamr-06-2018-0052

Badar, K., Aboramadan, M., & Plimmer, G. (2023). Despotic vs narcissistic leadership: Differences in their relationship to emotional exhaustion and turnover intentions. *International Journal of Conflict Management*, *34*(4), 818–837. https://doi.org/10.1108/IJCMA-12-2022-0210

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall. https://ci.nii.ac.jp/ncid/BA00746561

Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*(2), 248–287. https://doi.org/10.1016/0749-5978(91)90022-l

Bansal, S., Garg, N., & Singh, J. (2023). Weight-based teasing and depressive symptoms among Indian college students: Exploring the moderating effect of gratitude. *Kybernetes*, *53*, 2214. https://doi.org/10.1108/k-10-2022-1486

Bhalerao, H., & Kumar, S. (2015). Nonviolence at workplace—Scale development and validation. *Business Perspectives and Research*, *3*(1), 36–51. https://doi.org/10.1177/2278533714551863

Bhende, P., Mekoth, N., Ingalhalli, V., & Reddy, Y. (2020). Quality of work life and work–life balance. *Journal of Human Values*, *26*(3), 256–265. https://doi.org/10.1177/0971685820939380

Biswas, W., & Chakraborty, D. (2023). 'As you sow so you reap': A multi-case study on reshaping workforce well-being through positive sustainability at construction firms. *South Asian Journal of Business and Management Cases*, *12*(3), 333–351. https://doi.org/10.1177/22779779231203349

Caesens, G., Stinglhamber, F., Demoulin, S., De Wilde, M., & Mierop, A. (2019). Perceived organizational support and workplace conflict: The mediating role of failure-related trust. *Frontiers in Psychology*, *9*, 1–13. https://doi.org/10.3389/fpsyg.2018.02704

Chakraborty, D., Bhatnagar, S. B., Biswas, W., & Dash, G. (2022). The subtle art of effecting a four-day workweek to drive performance. *Management and Labour Studies*, *47*(3), 275–297. https://doi.org/10.1177/0258042x221082893

Chakraborty, D., & Biswas, W. (2021). Enlivening workplace climate through strategic human resource management initiatives: Unleashing its efficacy. *Business Perspectives and Research*, *9*(3), 427–445. https://doi.org/10.1177/2278533720983069

Chen, Z. X., & Francesco, A. M. (2003). The relationship between the three components of commitment and employee performance in China. *Journal of Vocational Behavior*, *62*(3), 490–510. https://doi.org/10.1016/s0001-8791(02)00064-7

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.

Cong, L., Li, C., Fan, J., & Nauta, M. M. (2015). Workplace conflict and absence/lateness: The moderating effect of core self-evaluation in China and the United States. *International Journal of Stress Management*, *22*(3), 243–269. https://doi.org/10.1037/a0039163

Cox, T., Kuk, G., & Leiter, M. P. (2017). *Burnout, health, work stress, and organizational healthiness*. (pp. 177–193). Routledge eBooks. https://doi.org/10.4324/9781315227979-14

Cvenkel, N. R. (2020). Striving for a violence "free" and healthy workplace: Insights from forestry Workers' perspectives. In *Developments in corporate governance and responsibility*. Emerald Publishing Limited. https://doi.org/10.1108/s2043-052320200000016009

D'Cruz, P., Noronha, E., Baillien, E., Catley, B., Harlos, K., Høgh, A., & Mikkelsen, E. G. (2021). Pathways of job-related negative behaviour. In *Handbooks of workplace bullying, emotional abuse and harassment*. Springer Singapore. https://doi.org/10.1007/978-981-13-0935-9

Deshpande, R., & Chakraborty, D. (2023). Impact of work stressors on career commitment – A study of Indian IT sector employees. *Prabandhan : Indian Journal of Management*, *16*(11), 27. https://doi.org/10.17010/pijom/2023/v16i11/173214

Donohue, W. A., Cai, D. A., Fink, E. L., & Umana, E. A. (2023). Conflict, communication and team collaboration: A comparison of Nigerians and US Americans. *International Journal of Conflict Management*, *34*(4), 649–667. https://doi.org/10.1108/IJCMA-07-2022-0123

Dukhaykh, S. (2023). Exploring the relationship between work–family conflict, family–work conflict and job embeddedness: Examining the mediating role of emotional exhaustion. *Psychology Research and Behavior Management*, *16*, 4859–4868. https://doi.org/10.2147/prbm.s429283

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Gaji, A. A., Madaki, A. A., Bello, M. S., & Hussaini, I. (2021). Antecedents of perceived organisational politics and psychological withdrawal. *East African Scholars Journal of Economics, Business and Management*, *4*(7), 138–145. https://doi.org/10.36349/easjebm.2021.v04i07.00

Garg, N., & Saini, D. (2023). Workplace spirituality to cultivate innovative work behaviour: The moderating role of perceived working conditions. *International Journal of Sociology and Social Policy*, *44*, 42. https://doi.org/10.1108/ijssp-03-2023-0050

George, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference, 11.0 update*. Allyn & Bacon.

Guenter, H., Van Emmerik, H., Schreurs, B., Kuypers, T., Van Iterson, A., & Notelaers, G. (2016). When task conflict becomes personal. *Small Group Research*, *47*(5), 569–604. https://doi.org/10.1177/1046496416667816

Gürlek, M. (2021). Workplace ostracism, Syrian migrant workers' counterproductive work behaviors, and acculturation: Evidence from Turkey. *Journal of Hospitality and Tourism Management*, *46*, 336–346. https://doi.org/10.1016/j.jhtm.2021.01.012

Hamouche, S., Koritos, C., & Papastathopoulos, A. (2023). Quiet quitting: Relationship with other concepts and implications for tourism and hospitality. *International Journal of Contemporary Hospitality Management*, *35*, 4297–4312. https://doi.org/10.1108/ijchm-11-2022-1362

Hamre, K. V., Fauske, M. R., Reknes, I., Nielsen, M. B., Gjerstad, J., & Einarsen, S. (2021). Preventing and neutralizing the escalation of workplace bullying: The role of conflict management climate. *International Journal of Bullying Prevention*, *4*(4), 255–265. https://doi.org/10.1007/s42380-021-00100-y

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *Communication Methods and Measures*, *14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

J, A. (2014). Determinants of employee engagement and their impact on employee performance. *∼the œInternational Journal of Productivity and Performance Management/International Journal of Productivity and Performance Management*, *63*(3), 308–323. https://doi.org/10.1108/ijppm-01-2013-0008

Jehn, K. A. (1995). A multimethod examination of the benefits and detriments of intragroup conflict. *Administrative Science Quarterly*, *40*, 256–282.

Khan, A. G., Li, Y., Akram, Z., & Akram, U. (2021). Does bad gossiping trigger for targets to hide knowledge in morally disengaged? New multi-level insights of team relational conflict. *Journal of Knowledge Management*, *26*(9), 2370–2394. https://doi.org/10.1108/jkm-08-2021-0609

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Lata, K., & Garg, N. (2023). Predicting nonviolent work behaviour among employees using machine learning techniques. *International Journal of Conflict Management*, *34*, 931–944. https://doi.org/10.1108/ijcma-04-2023-0074

Leiß, T. V., & Rausch, A. (2022). How personality, emotions and situational characteristics affect learning from social interactions in the workplace. *Vocations and Learning*, *16*(1), 73–97. https://doi.org/10.1007/s12186-022-09303-w

Li, M., Xu, X., & Kwan, H. K. (2021). Consequences of workplace ostracism: A meta-analytic review. *Frontiers in Psychology*, *12*, 1–14. https://doi.org/10.3389/fpsyg.2021.641302

Liao, Y., Liao, H., & Yen, H. R. (2022). The impact of role conflict on frontline employees' adaptive service behavior: The moderation effect of role ambiguity: An abstract. In *Springer eBooks* (pp. 515–516). https://doi.org/10.1007/978-3-030-95346-1_168

Lu, Y., Hu, X., Xiaoliang, H., Zhuang, X., Guo, P., Feng, L., Hu, W., Chen, L. Q., Zou, H., & Hao, Y. (2017). The relationship between job satisfaction, work stress, work–family conflict, and turnover intention among physicians in Guangdong, China: A cross-sectional study. *BMJ Open*, *7*(5), e014894. https://doi.org/10.1136/bmjopen-2016-014894

Mahipalan, M., & Garg, N. (2024). Does workplace toxicity undermine psychological capital (PsyCap) of the employees? Exploring the moderating role of gratitude. *International Journal of Organizational Analysis*, *32*(3), 476–503. https://doi.org/10.1108/IJOA-12-2022-3543

Meier, L. L., Gross, S., Spector, P. E., & Semmer, N. K. (2013). Relationship and task conflict at work: Interactive short-term effects on angry mood and somatic complaints. *Journal of Occupational Health Psychology*, *18*(2), 144–156. https://doi.org/10.1037/a0032090

Parayitam, S., Usman, S. A., Namasivaayam, R. R., & Naina, M. S. (2020). Knowledge management and emotional exhaustion as moderators in the relationship between role conflict and organizational performance: Evidence from India. *Journal of Knowledge Management*, *25*(6), 1456–1485. https://doi.org/10.1108/jkm-03-2020-0184

Pereira, V., Hadjielias, E., Christofi, M., & Vrontis, D. (2023). A systematic literature review on the impact of artificial intelligence on workplace outcomes: A multi-process perspective. *Human Resource Management Review*, *33*(1), 100857. https://doi.org/10.1016/j.hrmr.2021.100857

Podsakoff, P. M., MacKenzie, S. B., Lee, J., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Pradhan, R. K., & Jena, L. K. (2016). Employee performance at workplace: Conceptual model and empirical validation. *Business Perspectives and Research*, *5*(1), 69–85. https://doi.org/10.1177/2278533716671630

Sarkar, A., & Garg, N. (2020). "Peaceful workplace" only a myth? *International Journal of Conflict Management*, *31*(5), 709–728. https://doi.org/10.1108/ijcma-11-2019-0217

Sarkar, A., & Garg, N. (2021). A multilevel model exploring the relationships between individual spirituality, spiritual climate, and nonviolent work behavior. *Conflict Resolution Quarterly*, *38*(3), 209–229. https://doi.org/10.1002/crq.21298

Sarkar, A., & Palo, S. (2019). *Nonviolence behaviour in the workplace: Myth or reality?* (pp. 65–81). Palgrave Studies in Indian Management. https://doi.org/10.1007/978-3-030-13984-1_5

Sawyer, K., Thoroughgood, C. N., Stillwell, E. E., Duffy, M., Scott, K., & Adair, E. C. (2022). Being present and thankful: A multi-study investigation of mindfulness, gratitude, and employee helping behavior. *Journal of Applied Psychology*, *107*(2), 240–262. https://doi.org/10.1037/apl0000903

Serenko, A. (2023). The human capital management perspective on quiet quitting: Recommendations for employees, managers, and national policymakers. *Journal of Knowledge Management*, *28*, 27–43. https://doi.org/10.1108/jkm-10-2022-0792

Shah, M., & Hashmi, M. S. (2019). Relationship between organizational culture and knowledge hiding in software industry: Mediating role of workplace ostracism and workplace incivility. https://hdl.handle.net/10419/214259

Shin, Y., Hur, W., Park, K., & Hwang, H. (2020). How Managers' job crafting reduces turnover intention: The mediating roles of role ambiguity and emotional exhaustion. *International Journal of Environmental Research and Public Health*, *17*(11), 3972. https://doi.org/10.3390/ijerph17113972

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Tapper, J. (2022). Quiet quitting: Why doing the bare minimum at work has gone global. *The Guardian* http://www.theguardian.com/money/2022/aug/06/quiet-quitting-why-doing-the-bare-minimum-at-work-has-gone-global

Ter Hoeven, C. L., & Van Zoonen, W. (2020). Helping others and feeling engaged in the context of workplace flexibility: The importance of communication control. *International Journal of Business Communication*, *60*(1), 62–83. https://doi.org/10.1177/2329488419898799

Tu, Y., Sulistiawan, J., Ekowati, D., & Rizaldy, H. (2022). Work-family conflict and salespeople deviant behavior: The mediating role of job stress. *Heliyon*, *8*(10), e10881. https://doi.org/10.1016/j.heliyon.2022.e10881

Xia, Y., & Yang, Y. (2018). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428.

Xiong, J., & Chen, Y. (2018). Exploring the effect of coworkers' idiosyncratic deals on employees' work withdrawal behavior: Based on the perspective of equity theory. *Journal of Psychological Science*, *41*, 929–935.

How to cite this article: Bansal, S., & Garg, N. (2024). Quitting silently: A longitudinal research on the impact of workplace conflict and nonviolent work behavior. *Conflict Resolution Quarterly*, 1–18. https://doi.org/10.1002/crq.21444

15411508, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/crq.21444 by Delhi Technological University, Wiley Online Library on [17/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Electrical Engineering
https://doi.org/10.1007/s00202-024-02609-z

ORIGINAL PAPER

Real-time investigation of grid-interactive EV charger with two-stage bidirectional converter under wide voltage range scenarios

Gaurav Yadav[1] · **Mukhtiar Singh**[1]

Received: 10 May 2024 / Accepted: 8 July 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract
With the increasing popularity of electric vehicles, the grid-interactive electric vehicle chargers (GIEVC) are very vital in utilizing them for ancillary services to support grid through V2G mode of operation. Further, the front-end converter of charger may have to face several Power Quality (PQ) issues such as voltage sag, voltage swell, and grid distortions, which are very common in weak grid. Therefore, the proposed work is mainly focused on developing control algorithm for GIEVC in order to handle the PQ issues while providing seamless transition from G2V to V2G mode of operations. Since the proposed the charger is of Level 1 for domestic applications; the typical nonlinear domestic load is also compensated in order to avoid the exaggeration of weak grid issues. Here, the major challenge is in extracting the fundamental voltage component from non-ideal grid voltage for which a cascaded non-identical second-order generalized integrator has been utilized. The efficacy of these algorithms is studied in MATLAB/Simulink simulations and validated using the real-time hardware platform using OP4510 RTLAB provided by OPAL-rt.

Keywords Grid-interactive electric vehicle · Leaky least mean fourth · On-board bidirectional charger · Orthogonal signal generator · Wide voltage range

1 Introduction

In recent times, electric cars, or EVs, have become increasingly popular as an eco-friendly and high-performing substitute for conventional gasoline-powered automobiles. Due to developments in battery technology, an expanding infrastructure of charging stations, and rising consumer demand, EVs are now a practical choice for a large number of drivers worldwide. These vehicles, which utilize electric motors to drive the wheels and are propelled by electricity stored in a battery instead of gasoline, are revolutionizing the way we think about transportation [1, 2].

Furthermore, electric vehicles, which interact with the grid, are more prevalent in the global energy crisis and reduce greenhouse gases. But these grid-interactive electric vehicles have the challenges of keeping the power system overloaded when one or more EVs are connected to the grid at the same time [2–4]. In addition, these GIEV can transmit power from grid to vehicle (G2V), i.e., charging and when the load is not high, so they can deliver their surplus energy to the power grid [i.e., discharging via vehicle-to-grid (V2G)] [5, 6] during the peak hours (e.g., the evening in residential areas).

The previously described studies show that bidirectional chargers are practically feasible; yet, further research is required to understand how the grid interacts with chargers and loads linked at the point of common coupling (PCC) when there are various anomalies, such as sag and swell [7]. Furthermore, their duration may vary from several seconds to tens of milliseconds, contingent upon the dynamic properties of the electrical apparatuses linked to the network. A detailed network model is necessary in order to comprehend the impacts of GIEVC chargers on grid-connected systems and to put management mechanisms in place to lessen such effects. Because of this, a model of this kind should take switching dynamics into consideration. Small simulation time steps are necessary to account for the high switching frequency, which may reach up to 20 kHz, necessitating expensive and time-consuming calculations [8]. For grid-connected networks that use chargers and other different

✉ Gaurav Yadav
gaurav812786@gmail.com

Mukhtiar Singh
smukhtiar_79@yahoo.co.in

[1] Power Electronics Lab, Electrical Engineering Department, Delhi Technological University, Delhi 11042, India

Published online: 25 July 2024

types of equipment, this is not feasible. In power electronics, average modeling has been used to solve the aforementioned issue. By focusing on the behavior of the converter rather than the switching dynamics during a switching period, this modeling approach can improve simulation efficiency without compromising the ability to predict the steady-state and dynamic responses of the converter [9]. Therefore, it is a useful method for simulating EV chargers in grid-integrated systems. On the other hand, the dynamic response as discussed above comes into picture when there is increase or decrease in nonlinear load connected at the point of common coupling (PCC). Moreover, the grid side voltage disturbance at PCC is also a great issue and in anticipation to that the front end converter must be able to provide reactive power support (RPS). Therefore, maintaining requisite power quality standards at PCC within permissible limits is also one of the major challenges.

Therefore, PQ issues [10] mentioned above can be minimized by taking care of load current at the PCC. In addition, load current should be passed through the filters so as to result in the generation of the reference value for grid current [11]. An adaptive theory-based control utilizes a feedback mechanism and self-adjusts the internal parameters [12]. Some of the adaptive controls include LMS [13], VSSLMS [14], and RLS [15]. Regular shunt compensators had previously implemented LMS-based adaptive controllers because of their simplicity of implementation. Nevertheless, there is a greater steady-state inaccuracy and slower weight convergence with the traditional LMS. As a consequence, the LMS's performance is only passably good, and maybe better algorithms are needed. In this work, an LLMF control is presented. This approach is part of a novel family of stochastic gradient descent-based techniques and extracts the weight of load currents, which equates to the basic active component, and eliminates the mean fourth error. Through the reduction of detected current error, it offers a quick and precise method for estimating sinusoidal reference currents given distorted load currents. The application of this filter for shunt compensation is innovative and has not been reported in the literature [16]

In GIEV charger, the front-end converter (FEC) plays pivotal role in absorbing the grid side fluctuations while charging the battery and may provide ancillary services such as reactive power support and load harmonic compensation at PCC (as discussed above) if required [17]. Under normal grid conditions, there is no issue in FEC control. However, problem arises under non-ideal voltage conditions due to the presence of negative sequence and DC-offset components. To get rid of the negative sequence components in grid voltages, a Quadrature Signal Generator (QSG)-based two-channel method works well. Although QSG exhibits satisfactory performance in rejecting the negative sequence component as well as total harmonic distortion (THD) but have limitations in dealing with the DC-offset component [18]. Fourth-order QSGs (FO-QSGs) [19] provide complete dc-offset rejection and higher harmonics attenuation than second [20], third [21], and mixed (second and third) [22] QSGs. A SOSOGI [23] and a CSOGI [24] are two such FO-QSGs. Based on the system's settling time, the SOSOGI parameters are established. When choosing parameters, the fourth-order system in [25] is roughly equivalent to a lower-order system. Consequently, a sizable discrepancy exists between the predicted and actual settling periods for different combinations of damping factors and settling time. A CSOGI is created in [24] by cascading two SOGIs with comparable properties. Fourth-order harmonic attenuation is provided by the two cascaded SOGIs in CSOGI, while the first SOGI removes dc-offset from detected grid voltages. An over-damped cascaded SOGI might provide a more delayed reaction than a single SOGI, if the single SOGI-QSG damping factor is not adjusted correctly. This work proposes a CNISOGI filter to estimate the fundamental component (FC) from the polluted grid voltages to obtain improved compensation. Also, CNISOGI is the cascaded form of two SOGIs just like CSOGI, but differs in terms of having different or non-identical gains. Henceforth, it offers higher degree of freedom for achieving better settling time as compared to CSOGI [26]. Moreover, to achieve high reliability a real-time simulation of on-board bidirectional charger (OBBC), a CPU-FPGA-based real-time simulator, OP4510, from OPAL-RT Technologies is used.

The technical contributions involved in this microgrid are enumerated as follows.

- The complete equivalent model (CEM) simplifies OBBC by improving the simulation efficiency without compromising the accuracy of simulation performed [25].
- In order to compensate for imbalanced and distorted grid voltages, a CNISOGI filter is added for the first time to the OBBC. The grid voltages unit templates are generated by estimating the FC from the contaminated grid voltages. Therefore, even when the grid voltages exhibit anomalies, the grid currents obey a smooth, sinusoidal, and balanced pattern. It also computes the phase angle and frequency of the grid voltages. Thus, a smooth synchronization between the electrical grid and the OBBC is accomplished.
- Load compensation is achieved by using an LLMF adaptive filter, which calculates the fundamental active weight of load currents using a faster convergence rate with minimal weight oscillations. As a result, the supply current quality meets IEEE-519 requirements.

The structure of this document is as afterward. Section 1 provides a description of the literature review and recommended reading. The system description is discussed in Section 2. Controller design is discussed in Section 3.

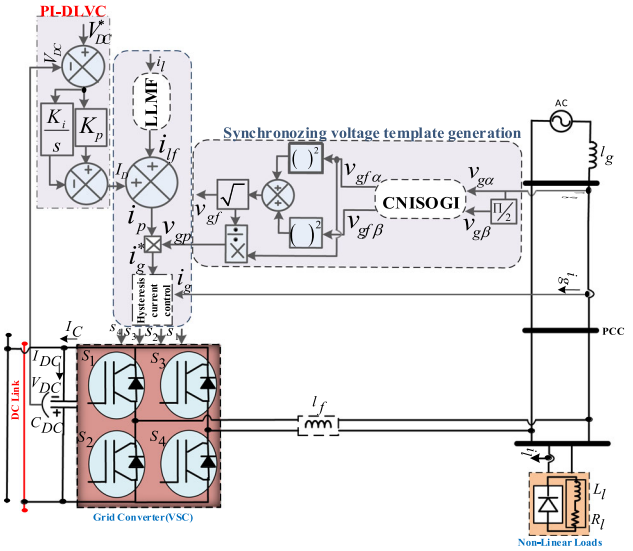


Fig. 1 Circuit Diagram of OBBC

Section 4 summarizes the suggested method. Section 5 summarizes about Results & discussion, & Section 6 about the conclusion.

2 System configuration

As shown in Fig. 1, the structure of OBBC under discussion consists of two conversion phases. The conversion of AC to DC is the responsibility of the first stage, and DC to DC is the domain of the second level. Additionally, the PCC has a nonlinear load linked. The first AC–DC converter in the recommended control configuration functions as a regulated rectifier, while the second phase uses a Bipolar Dual-Directional DC (BDDC) converter with two IGBT switches. After this setup, a mix of constant voltage (CV) and constant current (CC) methods are used to charge the lithium-ion battery. Source voltage (v_g), source inductance L_s, shunt R_l & L_l function as the load, and coupling inductance L_f are the different parameters and parts of the whole OBBC system. The switches (S_1, S_2, S_3, S_4) function as a bidirectional system due to the AC/DC inverter, particularly because of the DC link voltage's (V_{DC}) stiffness. DC/DC converter that performs buck-boost conversion, using Boost mode (S_6) for battery discharge and Buck mode (S_5) for charging batteries.

2.1 Modeling of single phase OBBC EV charger

Single phase OBBC EV chargers comprise of two stage AC/DC and DC/DC. In this section, firstly, the modeling of an AC/DC converter and secondly, a DC/DC converter are discussed.

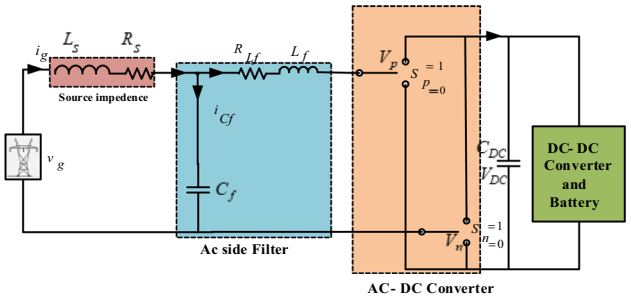


Fig. 2 Switching model of AC/DC converter

2.1.1 AC/DC converter

The switching architecture for a single-phase AC–DC converter is shown in Fig. 2. For each leg, a single pole double throw switch is used. The grid voltage as represented by v_g, the grid side inductor as L_s, the AC side filter inductance and capacitance as L_f and C_f, respectively, and the equivalent series resistance (ESR) of the filter inductance and capacitance as R_{Lf} and R_{Cf}. For switches S_1 and S_3, the on state is represented by $S_p = 1$ and $S_p = 0$, accordingly. where the on states of switches S_2 and S_4 are represented by the values $S_n = 1$ and $S_n = 0$, respectively. As a result, the grid current i_g becomes the DC side current I_{DC} and the variable V_{pn} equals V_{DC} when switches S_1 and S_4 are on and S_3 and S_2 are off.

$$v_{pn} = v_p - v_n \tag{1}$$

$$= S_p \times V_{DC} - S_n \times V_{DC} = (S_p - S_n)V_{DC} \tag{2}$$

$$= S_{pn} \times V_{DC} \tag{3}$$

where

$$S_{pn} = S_p - S_n$$

V_{DC}= Average DC link voltage.
And

$$I_{DC} = (S_p - S_n)i_{Lf} = S_{pn} \times i_{Lf} \tag{4}$$

Now applying KVL in outer loop and neglecting capacitor ESR,

$$v_g = v_{LS} + v_{Cf} \tag{5}$$

$$v_g = L_s \frac{di_g}{dt} + v_{Cf} \tag{6}$$

$$\frac{di_g}{dt} = \frac{v_g}{L_s} - \frac{v_{Cf}}{L_s} \tag{7}$$

Now applying KVL in the inner loop,

$$v_{Cf} = v_{Lf} + i_{Lf} \times R_{Lf} + v_{pn}$$

$$v_{\mathrm{Cf}} = L_{\mathrm{f}}\frac{\mathrm{d}i_{\mathrm{Lf}}}{\mathrm{d}t} + i_{\mathrm{Lf}} \times R_{\mathrm{Lf}} + v_{\mathrm{pn}} \tag{8}$$

$$\frac{\mathrm{d}i_{\mathrm{Lf}}}{\mathrm{d}t} = \frac{v_{\mathrm{Cf}}}{L_{\mathrm{f}}} - \frac{i_{\mathrm{Lf}} \times R_{\mathrm{Lf}}}{L_{\mathrm{f}}} - \frac{v_{\mathrm{pn}}}{L_{\mathrm{f}}} \tag{9}$$

where

v_{Lf}= AC side Inductor filter voltage.
i_{Lf}= AC side Inductor filter current.
R_{Lf}= Equivalent Series resistance of filter inductor.
L_{f}= AC side filter Inductance.
Now applying the KCL at the junction,

$$i_{\mathrm{g}} = i_{\mathrm{Lf}} + i_{\mathrm{Cf}}$$

$$i_{\mathrm{g}} = i_{\mathrm{Lf}} + C_{\mathrm{f}}\frac{\mathrm{d}v_{\mathrm{Cf}}}{\mathrm{d}t} \tag{10}$$

$$\frac{\mathrm{d}v_{Cf}}{\mathrm{d}t} = \frac{i_{\mathrm{g}}}{C_{\mathrm{f}}} - \frac{i_{\mathrm{Lf}}}{C_{\mathrm{f}}} \tag{11}$$

where

i_{g} = Grid Current.
i_{Lf} = AC side Inductor filter current.
i_{Cf} = AC side Capacitor filter current.
C_{f} = AC side filter Capacitance.

2.1.2 Average model

Using one switching period as the switching model, the averaging operator is used, the system average model is acquired. For a time-varying variable y, the averaging operator is illustrated as (12).

$$\tilde{y} = \frac{1}{T}\int_0^T y(\tau)\mathrm{d}\tau \tag{12}$$

where

$\tilde{y}$= Average Value.
$y(\tau)$ = Time-varying Variable.
T= Switching Period.
Applying the averaging operator to above derived state-space Eqns (5), (7), (9).

$$\frac{\mathrm{d}\tilde{i}_{\mathrm{g}}}{\mathrm{d}t} = \frac{\tilde{v}_{\mathrm{g}}}{L_{\mathrm{s}}} - \frac{\tilde{v}_{\mathrm{Cf}}}{L_{\mathrm{s}}} \tag{13}$$

$$\frac{\mathrm{d}\tilde{i}_{\mathrm{Lf}}}{\mathrm{d}t} = \frac{\tilde{v}_{\mathrm{Cf}}}{L_{\mathrm{f}}} - \frac{\tilde{i}_{\mathrm{Lf}} \times R_{\mathrm{Lf}}}{L_{\mathrm{Lf}}} - \frac{d_{\mathrm{pn}} \times V_{\mathrm{DC}}}{L_{\mathrm{f}}} \tag{14}$$

$$\frac{\mathrm{d}\tilde{v}_{Cf}}{\mathrm{d}t} = \frac{\tilde{i}_{\mathrm{g}}}{C_{\mathrm{f}}} - \frac{\tilde{i}_{\mathrm{Lf}}}{C_{\mathrm{f}}} \tag{15}$$

where

d_{pn} = The average value of S_{pn}

2.1.3 Small signal model

The steady-state values of the average model are combined with small AC changes that are overlaid to create the small signal model. The resulting Eqns 11, 12, and 14 for small signals are

$$\frac{\mathrm{d}\tilde{\tilde{i}}_{\mathrm{g}}}{\mathrm{d}t} = \frac{\tilde{\tilde{v}}_{\mathrm{g}}}{L_{\mathrm{s}}} - \frac{\tilde{\tilde{v}}_{\mathrm{Cf}}}{L_{\mathrm{s}}} \tag{16}$$

$$\frac{\mathrm{d}\tilde{\tilde{i}}_{Lf}}{\mathrm{d}t} = \frac{\tilde{\tilde{v}}_{\mathrm{Cf}}}{L_{\mathrm{f}}} - \frac{\tilde{\tilde{i}}_{\mathrm{Lf}} \times R_{\mathrm{Lf}}}{L_{\mathrm{Lf}}} - \frac{\tilde{d}_{\mathrm{pn}} \times V_{\mathrm{DC}}}{L_{\mathrm{f}}} \tag{17}$$

$$\frac{\mathrm{d}\tilde{\tilde{v}}_{\mathrm{Cf}}}{\mathrm{d}t} = \frac{\tilde{\tilde{i}}_{\mathrm{g}}}{C_{\mathrm{f}}} - \frac{\tilde{\tilde{i}}_{\mathrm{Lf}}}{C_{\mathrm{f}}} \tag{18}$$

Taking Laplace transform of Eqns 16-18,

$$\tilde{\tilde{i}}_{\mathrm{g}} = \frac{\tilde{\tilde{v}}_{\mathrm{g}}}{sL_{\mathrm{s}}} - \frac{\tilde{\tilde{v}}_{\mathrm{Cf}}}{sL_{\mathrm{s}}} \tag{19}$$

$$s\tilde{\tilde{i}}_{\mathrm{Lf}} = \frac{\tilde{\tilde{v}}_{\mathrm{Cf}}}{L_{\mathrm{f}}} - \frac{\tilde{\tilde{i}}_{\mathrm{Lf}} \times R_{\mathrm{Lf}}}{L_{\mathrm{f}}} - \frac{\tilde{d}_{\mathrm{pn}} \times V_{\mathrm{DC}}}{L_{\mathrm{f}}} \tag{20}$$

$$s\tilde{\tilde{v}}_{\mathrm{Cf}} = \frac{\tilde{\tilde{i}}_{\mathrm{g}}}{C_{\mathrm{f}}} - \frac{\tilde{\tilde{i}}_{\mathrm{Lf}}}{C_{\mathrm{f}}} \tag{21}$$

Putting Eqn (19) in Eqn (21)

$$S\tilde{\tilde{v}}_{\mathrm{Cf}} = \frac{1}{C_{\mathrm{f}}}\left(\frac{\tilde{\tilde{v}}_{\mathrm{g}}}{sL_{\mathrm{s}}} - \frac{\tilde{\tilde{v}}_{\mathrm{Cf}}}{sL_{\mathrm{s}}}\right) - \frac{\tilde{\tilde{i}}_{\mathrm{Lf}}}{C_{\mathrm{f}}} \tag{22}$$

$$\tilde{\tilde{v}}_{\mathrm{Cf}} = \frac{\tilde{\tilde{v}}_{\mathrm{g}}}{s^2L_{\mathrm{s}}C_{\mathrm{f}}+1} - \frac{\tilde{\tilde{i}}_{\mathrm{Lf}} \times sL_{\mathrm{s}}}{s^2L_{\mathrm{s}}C_{\mathrm{f}}+1} \tag{23}$$

Putting Eqn (23) in Eqn (20)

$$\tilde{\tilde{i}}_{\mathrm{Lf}} = \left(\frac{\tilde{\tilde{v}}_{\mathrm{g}}}{s^3L_{\mathrm{f}}L_{\mathrm{s}}C_{\mathrm{f}} + s^2L_{\mathrm{s}}C_{\mathrm{f}}R_{\mathrm{Lf}} + s(L_{\mathrm{f}}+L_{\mathrm{s}}) + R_{\mathrm{Lf}}}\right) - \left(\tilde{d}_{\mathrm{pn}} \times V_{\mathrm{DC}}\left(\frac{s^2L_sC_{\mathrm{f}}+1}{s^3L_{\mathrm{f}}L_{\mathrm{s}}C_{\mathrm{f}} + s^2L_{\mathrm{s}}C_{\mathrm{f}}R_{\mathrm{Lf}} + s(L_{\mathrm{f}}+L_{\mathrm{s}}) + R_{\mathrm{Lf}}}\right)\right) \tag{24}$$

Let us assume the ideal grid, then the final relation is as follows:

$$\tilde{\tilde{i}}_{\mathrm{Lf}} = -\tilde{d}_{\mathrm{pn}} \times V_{\mathrm{DC}} \times \left(\frac{s^2L_{\mathrm{s}}C_{\mathrm{f}}+1}{s^3L_{\mathrm{f}}L_{\mathrm{s}}C_{\mathrm{f}} + s^2L_{\mathrm{s}}C_{\mathrm{f}}R_{\mathrm{Lf}} + s\,(L_{\mathrm{f}}+L_{\mathrm{s}}) + R_{\mathrm{Lf}}}\right) \tag{25}$$

As it is known, at higher frequencies, the above equation results in (26).

$$\tilde{\tilde{i}}_{\mathrm{Lf}} = \frac{-\tilde{d}_{\mathrm{pn}} \times V_{\mathrm{DC}}}{sL_{\mathrm{f}} + R_{\mathrm{Lf}}} \tag{26}$$

From Eq. (26), it shows that $\tilde{d}_{\mathrm{pn}}$ increases with the decrement in $\tilde{\tilde{i}}_{\mathrm{Lf}}$ because of −ive sign. Therefore, the plant function is as follows in Fig. 6.

$$G_{P_{\mathrm{ac}}} = V_{\mathrm{DC}} \times \left(\frac{1}{sL_{\mathrm{f}} + R_{\mathrm{Lf}}}\right) \tag{27}$$

2.2 DC/DC converter

A pair of logical modes for the bidirectional buck-boost converter may be identified assuming complimentary switch control signals for S_5 and S_6 as shown in Fig. 10:

Topological mode 1 In this case, S_5 is ON and S_6 is OFF, yielding:

$$\frac{\mathrm{d}I_{L_{\mathrm{f}}}}{\mathrm{d}t} = \frac{1}{L_{\mathrm{f}}}V_{\mathrm{DC}} - \frac{1}{L_{\mathrm{f}}}V_b \tag{28}$$

$$\frac{\mathrm{d}V_{\mathrm{DC}}}{\mathrm{d}t} = \frac{1}{C_{\mathrm{DC}}}I_{\mathrm{DC}} - \frac{1}{C_{\mathrm{DC}}}I_{\mathrm{Lf}} \tag{29}$$

$$\frac{\mathrm{d}V_b}{\mathrm{d}t} = \frac{1}{C_{\mathrm{f}}}I_{\mathrm{Lf}} - \frac{1}{C_{\mathrm{f}}}I_b \tag{30}$$

Topological mode 2 In this case, S_5 is OFF and S_6 is ON, which leads to:

$$\frac{\mathrm{d}I_{L_{\mathrm{f}}}}{\mathrm{d}t} = -\frac{1}{L_{\mathrm{f}}}V_b \tag{31}$$

$$\frac{\mathrm{d}V_{\mathrm{DC}}}{\mathrm{d}t} = \frac{1}{C_{\mathrm{DC}}}I_{\mathrm{DC}} \tag{32}$$

$$\frac{\mathrm{d}V_b}{\mathrm{d}t} = \frac{1}{C_{\mathrm{f}}}I_{\mathrm{Lf}} - \frac{1}{C_{\mathrm{f}}}I_b \tag{33}$$

The following formulas may be obtained by averaging the first-order differential Eqs. (28–33) over a switching period; where D_S is the duty cycle for S_5 in Fig. 2:

$$\left.\frac{\mathrm{d}I_{L_{\mathrm{f}}}}{\mathrm{d}t}\right|_{\mathrm{av}} = D_{S_5}\left(\frac{1}{L_{\mathrm{f}}}V_{\mathrm{DC}} - \frac{1}{L_{\mathrm{f}}}V_b\right) + \left(1 - D_{S_5}\right)\left(-\frac{1}{L_{\mathrm{f}}}V_b\right) \tag{34}$$

$$= D_{S_5}\frac{V_{\mathrm{DC}}}{L_{\mathrm{f}}} - \frac{V_b}{L_{\mathrm{f}}} \tag{35}$$

$$\left.\frac{\mathrm{d}V_{\mathrm{DC}}}{\mathrm{d}t}\right|_{\mathrm{av}} = D_{S_5}\left(\frac{1}{C_{\mathrm{DC}}}I_{\mathrm{DC}} - \frac{1}{C_{\mathrm{DC}}}I_{L_{\mathrm{f}}}\right) + \left(1 - D_{S_5}\right)\left(\frac{1}{C_{\mathrm{DC}}}I_{\mathrm{DC}}\right) \tag{36}$$

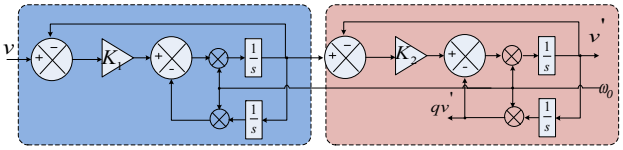


Fig. 3 Block diagram of suggested CNISOGI-QSG

$$= \frac{1}{C_{\mathrm{dc}}}I_{\mathrm{DC}} - D_{S_5}\left(-\frac{I_{L_{\mathrm{f}}}}{C_{\mathrm{DC}}}\right) \tag{37}$$

$$\left.\frac{\mathrm{d}V_b}{\mathrm{d}t}\right|_{\mathrm{av}} = \left(\frac{1}{C_{\mathrm{f}}}I_{L_{\mathrm{f}}} - \frac{1}{C_{\mathrm{f}}}I_b\right) \tag{38}$$

3 Control techniques

In context of control strategies, firstly, AC/DC converter control is discussed, followed by DC/DC control.

3.1 VSC switching strategy

Furthermore, Fig. 1 also states the switching strategy for VSC and the control steps are as follows:

3.1.1 Generation of unit templates

For the generation of the unit template of grid voltage (v_{g}) it is translated into $\alpha\beta$ -frame by simply taking grid voltage as $v_{g\alpha}$ and the same providing through transportation delay of 90° as $v_{\mathrm{g}\beta}$. In addition, it is processed through CNISOGI filter to remove the DC offset and harmonics components to get the fundamental $v_{\mathrm{g}f(\alpha\beta)}$.

The construction of a CNISOGI filter is given in Fig. 3. Additionally, the transfer function of the CNISOGI filter in-phase element (v') and quadrature-phase element (qv') to the input (v) is provided in (39) and (40) and may be derived in Fig. 3.

$$G_1(s) = \frac{v'(s)}{v(s)} = \left(\frac{K_1\omega_o s}{s^2 + K_1\omega_o s + \omega_o^2}\right)\left(\frac{K_2\omega_o s}{s^2 + K_2\omega_o s + \omega_o^2}\right) \tag{39}$$

$$H_1(s) = \frac{qv'(s)}{v(s)} = \left(\frac{K_1\omega_o s}{s^2 + K_1\omega_o s + \omega_o^2}\right)\left(\frac{K_2\omega_o^2 s}{s^2 + K_2\omega_o s + \omega_o^2}\right) \tag{40}$$

Figure 4 shows the pole-zero map of the CNISOGI filter. Furthermore, it can be seen that all poles lie on the left-hand side of the imaginary axis which states the stability of the system.

Figure 5 relates the Bode Diagrams of the in-phase com-

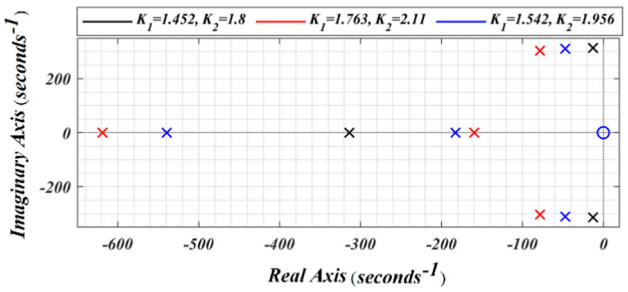


Fig. 4 Polar plot for CNISOGI

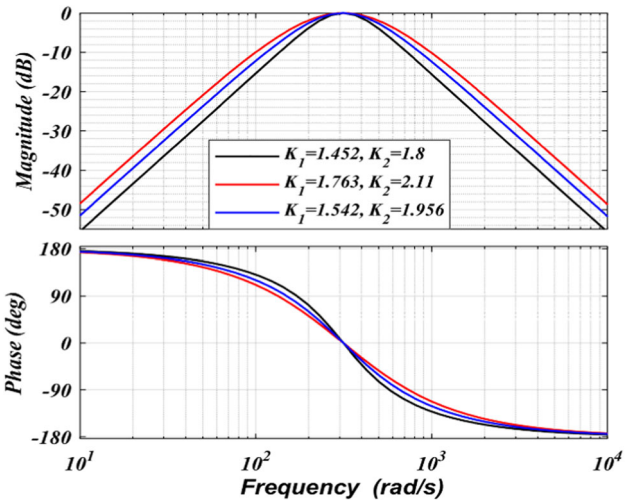


Fig. 5 Bode Diagram for CNISOGI

ponents of the CNISOGIs-QSG for the different values of K_1 & K_2. It is noticed that the CNISOGIs-QSG delivers stronger attenuation to harmonic components.

Furthermore, the $v_{gf\alpha}$ along with its phase-shifted quadrature component $v_{gf\beta}$ are conceded through a peak approximation block to calculate v_{gf}. Additionally, calculated v_{gf} is divided by $v_{gf\alpha}$ to obtain unit template v_{gp} as given in Eqs. (41), (42).

$$v_{gf} = \sqrt{\left(v_{gf\alpha}^2 + v_{gf\beta}^2\right)} \tag{41}$$

$$v_{gp} = \frac{\left(v_{gf\alpha}\right)}{((v_{gf\alpha})^2 + (v_{gf\beta})^2)^{0.5}} \tag{42}$$

Henceforth, to validate the effectiveness of the proposed method in extracting the in-phase and quadrature components out of a non-ideal voltage signal, an input supply comprising $5th$, $7th$, $11th$, and $13th$ harmonics and a dc offset, with an overall THD of 15%, has been considered, as shown in Fig. 6. The extracted values of in-phase and quadrature components utilizing and *CNISOGI* are shown in Fig. 7. On analyzing these waveforms, it can be observed

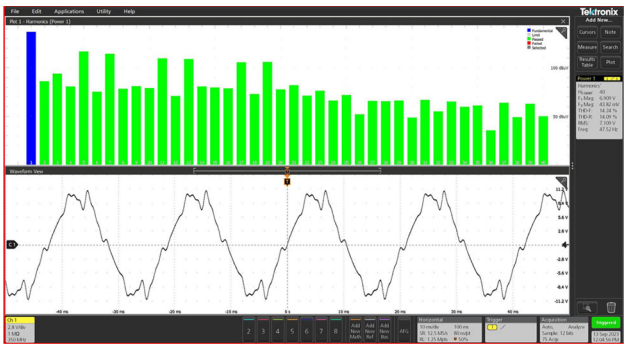


Fig. 6 Grid voltage comprising 5th, 7th, 11th, and 13th harmonics and a DC offset, with an overall THD of 15%

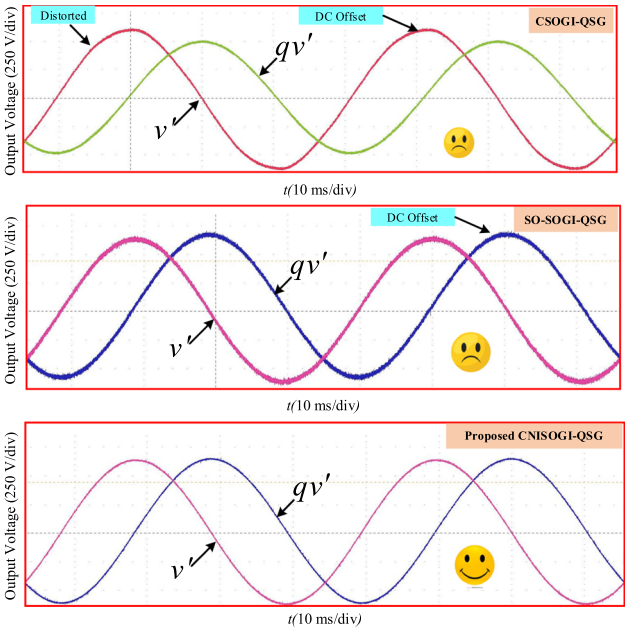


Fig. 7 Experimental output waveforms, spectrums of v' & spectrums of qv' for different *QSG* w.r.t output waveform

that *CSOGI* and *SO-SOGI* are not immune to grid side disturbances as they introduce DC-Offset and ripples in extracted components. However, the *CNISOGIs* approach tackles the grid side THDs effortlessly and extracts the fundamental component without any DC offset and ripples. Henceforth, comparison in terms of computational burden of proposed method (CNISOGI) with the other methods, i.e., CSOGI and SO-SOGI was performed on the real-time simulator using *d*-Space-1104 s. So, the proposed filter takes 3.7365E−05 while CSOGI takes 3.8489E−05 and SO-SOGI takes 3.94496E−05 turnaround time on the sample time of 5E−05. Therefore, it can be concluded that CNISOGI is 3.008% and 5.579% efficient than CSOGI and SO-SOGI, respectively.

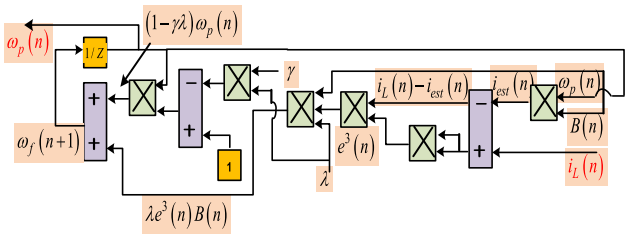


Fig. 8 Circuit Diagram of LLMF Control Architecture

3.1.2 Estimation of the active weight loss component of load current:

To calculate the switching loss within the shunt compensator V_{DC}^{*}, V_{DC} is passed via an LPF. Given the inaccuracy of the V_{DC} for feedback, a PI controller estimates the active component of DC-link voltage regulation and losses generated by the PI-DC link voltage controller (DLVC). A stable 400 V V_{DC} is maintained.

$$\omega_{\text{loss}}(n+1) = \omega_{\text{p}}(n) + [k_{\text{P}}\{V_{\text{dce}}(n+1) - V_{\text{dce}}(n)\} + k_{\text{i}}\{V_{\text{dce}}(n+1)\}] \tag{43}$$

where the constants for integral & proportional controller gains are represented by the letters k_{i} and k_{p}.

3.1.3 Generation of grid reference current & gating pulses:

About getting the pulse load current is fed to the LLMF filter as shown in Fig. 8 (for Eq. (59) explained in Section 5) which gives the $\omega_{\text{p}}(n) = i_{\text{lf}}$.

Furthermore, DC link loss component I_D and the fundamental load current component i_{lf} blend to prompt an overall active current reference (I_{p}) as given by Eq. (44).

$$I_{\text{p}} = I_D + I_{\text{lf}} \tag{44}$$

The active current reference I_{p} is multiplied with unit voltage template v_{gp} to provide the reference sinusoidal current corresponding to active power as given by Eq. (45).

$$i_{\text{g}}^{*} = I_{\text{p}} * v_{\text{gp}} \tag{45}$$

At last, these desired reference values are after comparison from actual grid current i_{g} is fed to Hysteresis current control (HCC) to compute gating signals GSC's (Grid side converter) IGBT switches.

Henceforth, Fig. 9 validates about the active fundamental weight (ω_{p}) and active current reference (I_{p}). As discussed earlier that ω_{p} is the filtered part of the load current generated by the LLMF.

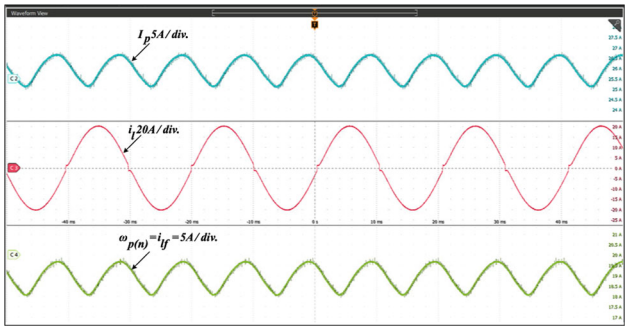


Fig. 9 Active fundamental weight (ω_p) and active current reference (I_p)

3.2 BDDC control in GCO

The primary goal of BDDC control is to continue charging or draining the battery in spite of grid-side interruptions. Here, the battery is charged using both constant voltage (CV) and constant current (CC). First, CC charging is used to manage battery charging. Once the battery voltage exceeds the allowed permissible voltage level, CV charging is introduced. This technique aids in quick battery charging without going over the charger's rated voltage and current limitations. The only thing left to do is build the battery reference current such that it includes CC/CV charging, as the DC-Link voltage is almost set at the rated voltage. Furthermore, since BDDC is a bidirectional system, its operation may be managed in both charging and discharging modes. Thus, in order to accomplish the intended goals, the controller must follow the reference current (I_b^{*}) for the battery, which may be developed based on the requirements. The BDDC's whole controller architecture is shown in Fig. 10.

4 Mathematical modeling of leaky least mean fourth algorithm

LLMF algorithm is a member of an entirely novel category of stochastic gradient descent-based algorithms that aims to reduce the mean fourth error, has become convex function for weight vector that is adaptable. Compared to the LMF method, the LLMF algorithm has a smaller steady-state error and quicker initial convergence, which is its main strength. As depicted in Fig. 11, N symbolizes the overall amount of matrix vectors.

The initial stages of interpolation involve identifying the interval index i that aligns with the input signal $t(n)$ and allocating the interpolation intervals, which are pivotal steps in the process. The control points for the look-up table (LUT) are given as follows:

$$A_N \in \Re^{Ax1} = [A_1 A_2 A_3 \ldots\ldots\ldots\ldots A_N]^T,$$

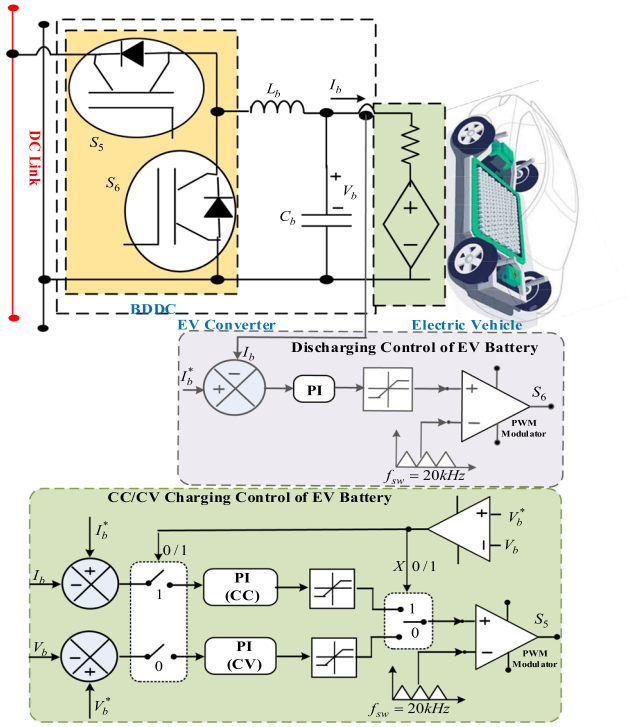


Fig. 10 BDDC control architecture

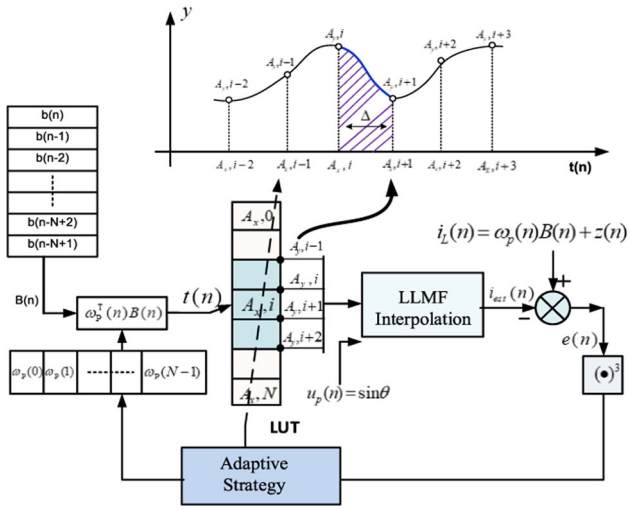


Fig. 11 Weight updating structure of LLMF adaptive algorithm

Where the ith control point denotes the A_i and $A_N \in \Re^{4x1} = A_{i-1}A_iA_{i+1}A_{i+2}]$ for the ith span. Additionally, by deducting the LLMF interpolation output out of the actual load current, $i_L(n)$, an error, $e(n)$, is derived. Once load dynamics are achieved, this adaptive progression is maintained until stable condition is attained.

Let us presume that the system receives the signal $B(n)$ together with a new weight coefficient vector.

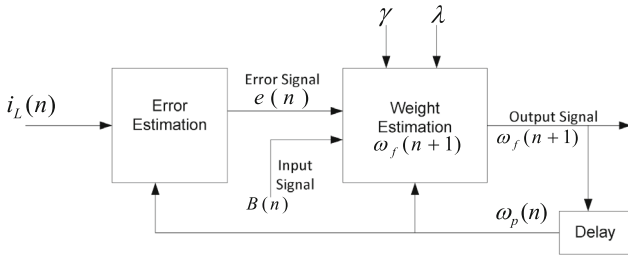


Fig. 12 Block diagram of LLMF control algorithm

$\omega_p(n) = \left[\omega_p(0)\omega_p(1)\omega_p(2)\ldots\omega_p(N-1)\right]^T$; and that the observed output signal $i_L(n)$ is given by (8)

$$i_L(n) = \omega_p(n) \times B(n) + z(n)$$

$$\mathrm{i}_L(n) = \omega_\mathrm{p}(n)B(n) + z(n) \text{ where}$$

$$B(n) = [b(n)b(n-1)....b(n-N+1)]^T$$

$z(n)$ represents the discovered noise, where is considered to be independent of $B(n)$, and shows the vector of the input signal $B(n)$

The objective is to use the input signal $B(n)$ & the measured output signal $i_L(n)$ to adaptively estimate the undefined channel coefficient vector $\omega_p(n)$. In Fig. 12, the LLMF block diagram is shown. Utilizing this block allows for the real-time estimation of the load current's basic weight and variations in weight. Comparing the LLMF control approach to the LMF method, it has been shown that the latter rapidly decreases error to zero.

In any AC system; the nonlinear load comprises the total of the fundamental and harmonic components, shown as;

$$i_L(t) = i_{L1}\sin(wt+\theta) + i_{L2}\sin(2\omega t+\theta) + \ldots\ldots\ldots\text{¥} \tag{46}$$

$$i_L(t) = i_{L1}\sin(wt+\theta) + \sum_{r=2}^{\text{¥}} i_{Lr}\sin(r\omega t+\theta_r) \tag{47}$$

where ω signifies the angular frequency of the fundamental harmonics, r indicates the order of harmonics, i_{L1} defines the fundamental harmonic, and i_{Lr} indicates the rth harmonics component of the current.

$$i_L(t) = i_{L1}[\sin wt\cos\theta + \sin\theta\cos\omega] + \sum_{r=2}^{\text{¥}} i_{Lr}[\sin r\omega\cos\theta_r + \cos r\omega t\sin\theta_r] \tag{48}$$

QUOTE $i_L(t) = i_{L1}[\sin\omega t\cos\theta + \sin\theta\cos\omega t] + \sum_{r=2}^{\infty} i_{Lr}[\sin r\omega t\cos\theta_r + \cos r\omega t\sin\theta_r]$

$i_{L1}\cos\theta$; $i_{L1}\sin\theta$; $i_{Lr}\cos\theta_r$; $i_{Lr}\sin\theta$ by $\omega_p(0)$; $\omega_p(1)$; $\omega_p(2)$; $\omega_p(3)$ and even harmonics are neglected

$$i_L(t) = \left[\omega_p(0)\sin\omega t + \omega_p(1)\cos\omega t\right] + \sum_{r=3,5\ldots}^{¥}\left[\omega_p(2)\sin r\omega t + \omega_p(3)\cos r\omega t\right] \quad (49)$$

In matrix form,

$$\left[\omega_p(0)\ \omega_p(1)\ \omega_p(2)\ \omega_p(3)\ ..\ ..\right]\begin{bmatrix}\sin\omega t\\ \cos\omega t\\ \sin 3\omega t\\ \cos 3\omega t\\ .\\ .\\ ¥\end{bmatrix} \quad (50)$$

where $B(n) = [sin\,\omega t cos\,\omega t sin3\,\omega t cos3\,\omega t.....\infty]^T$

$$\omega_p(n) = \left[\omega_p(0)\ \omega_p(1)\ \omega_p(2)\ \omega_p(3)\ ..\ ..\right] \quad (51)$$

Further, they used to calculate the load current presented as,

$$i_{\text{est}}(n) = \omega_p(n) \times B(n) \quad (52)$$

Let, Actual nonlinear load current and Rough current be $i_L(n)$ & $i_{est}(n)$, respectively, used to assess the error signal $e_p(n)$.

$$e(n) = i_L(n) - \omega_p(n) \times B(n) \quad (53)$$

With the LMF method, the cost function that has to be reduced is provided by

$$C_{\text{LMF}}(n) = 1/4 \times \left[e^4(n)\right] = 1/4 \times \left[\left\{i_L(n) - \omega_p(n)B(n)\right\}\right]^4 \quad (54)$$

The LLMF algorithm, together with the leakage parameter (γ), introduces the modified LMF cost function.

Consequently, the LLMF's minimal cost function is

$$C_{\text{LMF}}(n) = 1/4 \times \left[e^4(n)\right] + 2\gamma\left|\omega_p(n)\right|^2 \quad (55)$$

To get the partial derivative for cost function in relation to the weight vector is hence the definition of the gradient estimate.

$$\frac{\partial C(n)}{\partial \omega_p(n)} = \left[\left\{\gamma\omega_p(n) - e^3(n)B(n)\right\}\right] \quad (56)$$

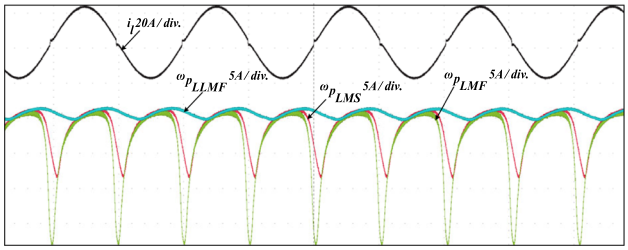


Fig. 13 Fundamental Active weight of load current

The gradient approach, sometimes referred to as the weight update equation of the LLMF, may be used iteratively to find the minimum of $C(n)$.

$$\omega_f(n+1) = \omega_p(n) - \lambda\frac{\partial C(n)}{\partial \omega_p(n)} \quad (57)$$

$$= \omega_p(n) - \lambda\left[\left\{\gamma\omega_p(n) - e^3(n)B(n)\right\}\right] \quad (58)$$

$$\omega_f(n+1) = \omega_p(n)(1-\lambda\gamma) + \lambda e^3(n)B(n) \quad (59)$$

where λ denotes the adaptation constant & γ is learning rate & its value ranges between 0 and 1.

Henceforth, preciseness of LLMF is shown in Fig. 13 where, for load current of 20A/div., active fundamental weight of different filter methods is compared and it can be observed among LMS and LMF, LLMF gives better result.

5 Result and discussion

5.1 Simulation validation

Using a variety of Simscape library components, the system is methodically designed and simulated in MATLAB/Simulink. A 230 v_{rms}, 50 Hz single-phase supply is used as the input for the OBBC FEC. Additionally, a nonlinear load that represents the normal home load has been linked to the PCC. Results from the simulation are available for two distinct periods. Figure 14 illustrates the controller's performance in the first phase while operating in the G2V and V2G modes. To enable UPF grid operation, the charger initially operates in G2V mode while meeting the reactive power needs of adjacent loads. The transition from G2V to V2G mode of operation occurs at 0.5 s. It is important to note that the transition between the two modes of operation is quite seamless. Furthermore, since the charger side is the sole source of reactive power, the grid always runs at UPF even when reactive power demand is at PCC. As the grid's present profile makes clear, the grid must only ever interact with active electricity. While the grid current operates in

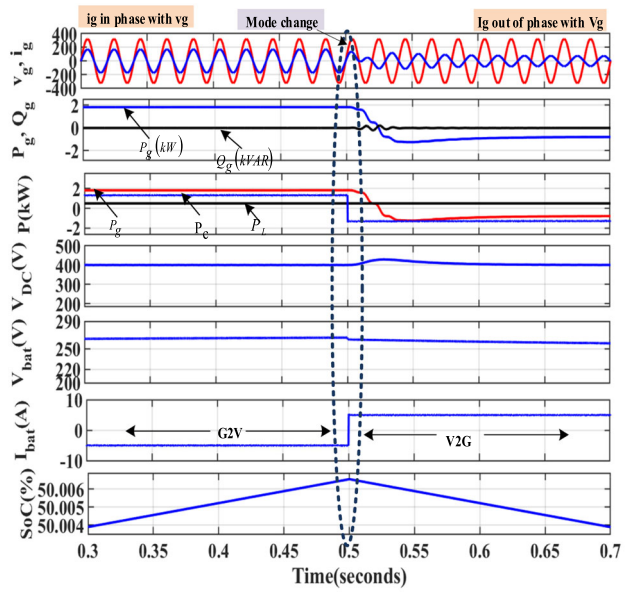


Fig. 14 Steady-state performance under normal operating conditions

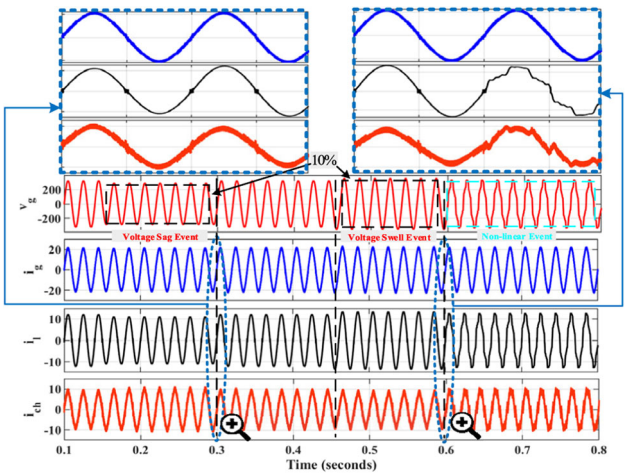


Fig. 15 Trace of Grid voltage (v_g), Grid current (i_g), load current (i_l), and FEC current (i_{ch})

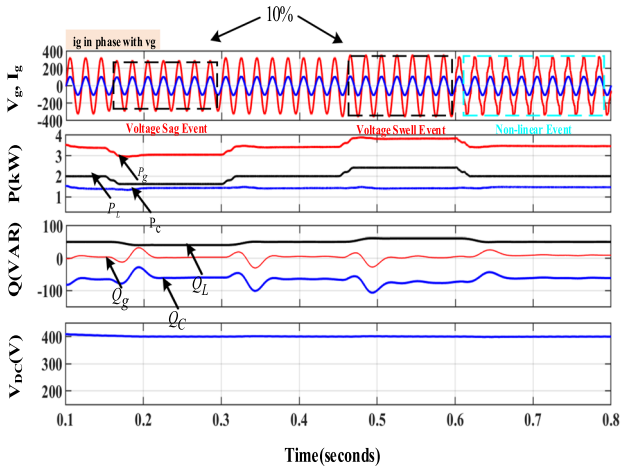


Fig. 16 Steady-state performance when harmonics are injected (weak grid operating condition)

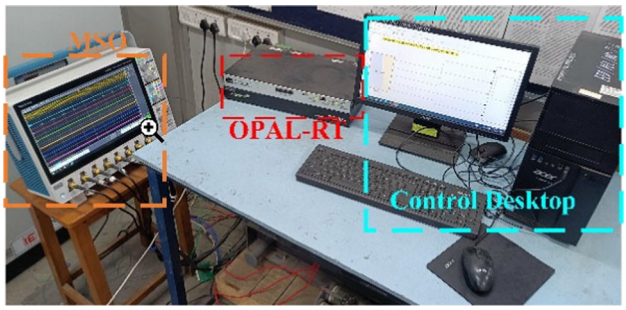


Fig. 17 Experimental Configuration

phase with the grid voltage in the G2V mode, it is out of phase with the grid voltage in the V2G mode.

During the second operating phase, various grid-side voltage disturbances are taken into consideration for system functioning. Here, as shown in Figs. 15 and 16, the voltage sag as well as swell of 10% with voltage THDs of 15% has been taken to examine the control performance of the charger. At first, as was previously said, the charger operates under typical grid circumstances. There is a voltage drop that occurs in 0.15–0.3 sec. The voltage returns to normal between 0.3 and 0.45. Finally, from 0.6 to 0.8, the nonlinear grid voltage with 15% THDs is taken into consideration. A voltage swell is inserted between 0.45 and 0.6. The waveforms for load current, FEC current, grid voltage, and grid current are shown in Fig. 15. Here, the FEC is set up to readily correct for both load harmonics and reactive power while concurrently charging the battery under various grid voltage disruptions. The grid voltage, current, active/reactive power of the FEC, load, and grid along the DC-Link voltage are all shown in Fig. 16, which further supports this claim. Here, grid voltage and grid current at UPF are always sinusoidal. This is further shown by the almost negligible reactive power on the grid side, which is a complement to the reactive power on the load from the charger.

5.2 Real-time HIL validation

The designed controller has been implemented into operation in real-time on a Hardware in Loop (C-HIL) configuration in order to fully evaluate its robustness and efficiency. The experimental setup, as shown graphically in Fig. 17, offers a thorough rundown of all the key parts that make up the HIL system. OPAL-RT ensures smooth integration and

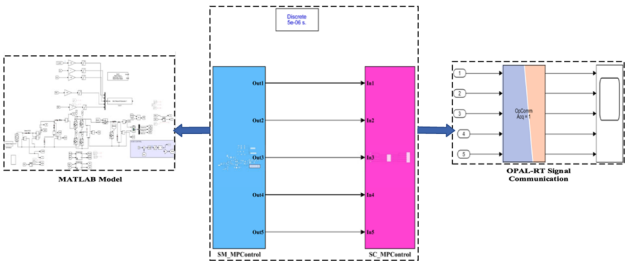


Fig. 18 OPAL-RT Simulator Model

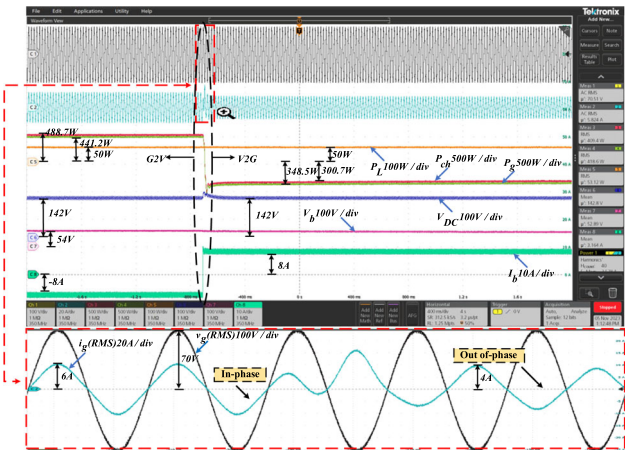


Fig. 19 Real-time *HIL* of Steady-state performance under normal operating conditions

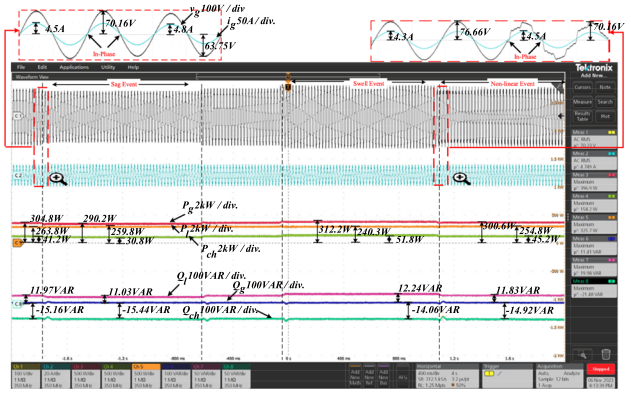


Fig. 20 Real-time validation of steady-state performance when harmonics are injected (weak grid operating condition)

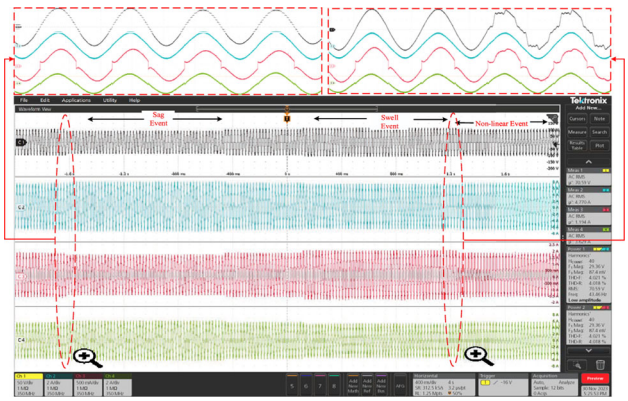


Fig. 21 Real-time *HIL* of Grid voltage (v_g), Grid current (i_g), load current (i_l), and FEC current (i_{ch})

accurate control by combining hardware and software synchronization options in a synergistic manner. In particular, the software synchronized mode is used for analysis in order to carefully assess the control method's effectiveness. The system performance is validated using MATLAB 2018a and the OP4510 model of RT-Lab 2019.2.3. Additionally, this part provides a comprehensive exhibition of the platform-obtained example test results depicted in Fig. 18, which serve to successfully reinforce the resilience and usefulness of the designed controller.

Figure 19 illustrates the Grid Voltage $v_g = 70\ v_{rms}$ at *100 V*/div. on Channel 1, and the real-time HIL results of the mode change from G2V to V2G. Channel 2 is devoted to Grid Current i_g. In charging mode (G2V), it draws a current of 6*A* (20*A*/div), and in discharging mode (G2V to V2G), it injects a current of 4A. The seamless transition from G2V to V2G with guaranteed UPF grid functioning while providing the load reactive demand from the FEC side alone is seen in the magnified version of v_g and i_g below. The power used by the load (50 W) and charger (440 W) in the presence of a nonlinear load connected at PCC may be determined to be provided by the grid (500 W) during G2V. After that, a reversal of power flow is seen as soon as the transition takes place. This is because the load (50 W) is receiving the required power from the charger (350 W), and the excess power (290 W) is being delivered into the grid, creating an out-of-phase grid current. Throughout all modes of operation, the DC-link voltage and battery voltage, which are 140 V at 100 V/div and 54 V at 100 V/div, respectively, are kept constant. Indicating the G2V and V2G modes of operation, the battery current may also follow the reference current change from − 8 to 8 A at 10 A/div.

Figs. 20, 21, and 22 show how the system might behave dynamically under non-ideal voltage settings. Initially, the optimal voltage is applied in this case. However, over time, voltage sag and swell of 10% and voltage THDs of 15% have been added to analyze the control performance. The charger is completely capable of responding to control instructions without being impacted by grid-side disruption, as can be shown by carefully examining the different waveforms. The battery current follows the reference current instruction with ease, the DC-link is almost always kept almost constant, and the FEC may concurrently charge and discharge the battery and correct for reactive power and load harmonics. The grid current at UPF consistently maintains a sinusoidal pattern

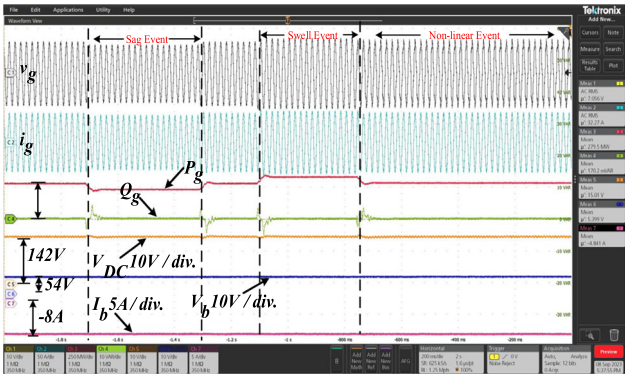


Fig. 22 Real-time validation of Steady-state performance when sag & swell in grid voltage

despite non-ideal voltage and nonlinear load at PCC, as can be seen from the magnified version of the grid voltage, grid current, load current, and FEC current shown in Fig. 21. Additionally, the reactive power profile of the grid, load, and FEC may be used to verify this fact. It shows that the UPF grid operates as shown in Figs. 20 and 22 when the FEC is able to fulfill the reactive power load requirement, keeping grid reactive power consistently around zero.

6 Conclusion

In this paper, a two-stage GIEVC charging system has been modeled, simulated, and validated through HIL in real-time. Furthermore, a fourth-order *CNISOGI* is realized and implemented. This unusual system effectively addresses irregularities in utility grid voltages by estimating synchronizing voltage template precisely. Moreover, it provides the sinusoidal nature of grid current although grid voltages are polluted. Also, LLMF filter approach is implemented for load compensation such as reactive power compensation and harmonics mitigation. The success of this technique is proven by *MATLAB/Simulink* simulations, specifically within the context of the V2G/G2V charging mode for *GIEVC* under voltage sag/swell and harmonics without affecting their normal operation. Also, to achieve high reliability, a real-time simulation of GIEVC, a real-time simulator, *OP4510*, from OPAL-RT Technologies is used. Additionally, the EV charger is easily able to compensate load harmonics while simultaneously regulating its output active and reactive power as per the set references.

Author contributions G.Y has drawn the required diagram and written the main manuscript. M.S has reviewed the manuscript.

Funding The authors did not receive any organization for the submitted work.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. https://www.investindia.gov.in/team-india-blogs/indias-ev-economy-future-automotive-transportation
2. https://www.india-briefing.com/news/indias-fame-ii-scheme-for-electric-vehicles-gets-80-bump-in-budget-allocation-27121.html/
3. Laha A, Yin B, Cheng Y, Cai LX, Wang Y (2019) Game theory based charging solution for networked electric vehicles: a location-aware approach. IEEE Trans Veh Technol 68(7):6352–6364. https://doi.org/10.1109/TVT.2019.2916475
4. Yadav G, Singh M (2023) Unveiling the superiority: comparative analysis of ANFIS, FOPID, and PI controllers in grid-connected EV systems for G2V/V2G applications, In *2023 International conference on electrical, electronics, communication and computers (ELEXCOM)*, Roorkee, India, pp 1–6, https://doi.org/10.1109/ELEXCOM58812.2023.10370196
5. Shah V, Payami S (2023) Integrated converter with G2V, V2G, and DC/V2V charging capabilities for switched reluctance motor drive-train based EV application. IEEE Trans Ind Appl 59(3):3837–3850. https://doi.org/10.1109/TIA.2023.3242636
6. Yadav G, Singh M (2023) Robust control design for grid-tied EV system with SOGI-based architecture in wide voltage range scenarios, In *2023 international conference on electrical, electronics, communication and computers (ELEXCOM)*, Roorkee, India, pp 1–6, https://doi.org/10.1109/ELEXCOM58812.2023.10370088
7. Saroha J, Singh M, Jain DK (2018) ANFIS-based add-on controller for unbalance voltage compensation in a low-voltage microgrid. IEEE Trans Ind Inf 14(12):5338–5345. https://doi.org/10.1109/TII.2018.2803748
8. Şahİn ME, Blaabjerg F, Sangwongwanİch A (2021) Modelling of supercapacitors based on simplified equivalent circuit. CPSS Trans Power Electron Appl 6(1):31–39. https://doi.org/10.24295/CPSSTPEA.2021.00003
9. Fan B, Wang X (2022) Equivalent circuit model of grid-forming converters with circular current limiter for transient stability analysis. IEEE Trans Power Syst 37(4):3141–3144. https://doi.org/10.1109/TPWRS.2022.3173160
10. Singh M, Khadkikar V, Chandra A, Varma RK (2011) Grid interconnection of renewable energy sources at the distribution level with power-quality improvement features. IEEE Trans Power Delivery 26(1):307–315. https://doi.org/10.1109/TPWRD.2010.2081384
11. Narayanan V, Singh B (2023) CNISOGI and VSSMCLMS filters for solving grid and load abnormalities for a microgrid with seamless mode transfer capability. IEEE Trans Ind Appl 59(4):4225–4237. https://doi.org/10.1109/TIA.2023.3259387
12. Diniz Paulo (2008) Adaptive filtering: algorithms and practical implementation. Springer, Boston. https://doi.org/10.1007/978-1-4614-4106-9
13. Batabyal T, Weller D, Kapur J, Acton ST (2023) Efficient learning of transform-domain LMS filter using graph Laplacian. IEEE Trans Neural Netw Learn Syst 34(10):7608–7620. https://doi.org/10.1109/TNNLS.2022.3144637
14. Shi J, Zhu H (2023) Control study for compensating rotor vibration of four-DOF Six-pole hybrid magnetic bearings based on variable step size LMS algorithm. IEEE J Emerg Select Top Power Electron 11(2):1616–1626. https://doi.org/10.1109/JESTPE.2022.3222656
15. Bhunia M, Subudhi B (2022) A self-tuning adaptive control scheme for a grid-connected three-phase PV system. IEEE J Emerg

Select Top Power Electron 10(5):5709–5716. https://doi.org/10.1109/JESTPE.2022.3162408
16. Khattak O-U-R, Zerguine A (2013) Leaky least mean fourth adaptive algorithm. IET Signal Process 7:134–145
17. Mishra MK, Lal VN (2023) A multiobjective control strategy for harmonic current mitigation with enhanced LVRT operation of a grid-tied PV system without PLL under abnormal grid conditions. IEEE J Emerg Select Top Power Electron 11(2):2164–2177. https://doi.org/10.1109/JESTPE.2022.3218888
18. Mohamadian S, Pairo H, Ghasemian A (2022) A straightforward quadrature signal generator for single-phase SOGI-PLL with low susceptibility to grid harmonics. IEEE Trans Industr Electron 69(7):6997–7007. https://doi.org/10.1109/TIE.2021.3095813
19. Mondal B, Karuppaswamy BA (2022) A new approach to fourth-order quadrature signal generation for a fast & noise-free PLL output under non-ideal grid voltage conditions. IEEE Access 10:38472–38482. https://doi.org/10.1109/ACCESS.2022.3165561
20. Sahoo A, Ravishankar J, Jones C (2021) Phase-locked loop independent second-order generalized integrator for single-phase grid synchronization. IEEE Trans Instrum Measure 70:1–9. https://doi.org/10.1109/TIM.2021.3104406
21. Tripathy P, Misra B, Nayak B (2022) Performance analysis of dual third order generalized integrator based phase locked loop for three phase distorted grid condition, In *2022 1st IEEE international conference on industrial electronics: developments & applications (ICIDeA)*, Bhubaneswar, India, pp 135–140, https://doi.org/10.1109/ICIDeA53933.2022.9970148
22. Zhang C, Zhao X, Wang X, Chai X, Zhang Z, Guo X (2018) A grid synchronization PLL method based on mixed second- and third-order generalized integrator for DC offset elimination and frequency adaptability. IEEE J Emerg Select Top Power Electron 6(3):1517–1526. https://doi.org/10.1109/JESTPE.2018.2810499
23. Gou L, Wang C, You X, Zhou M, Dong S (2022) IPMSM sensorless control for zero- and low-speed regions under low switching frequency condition based on fundamental model. IEEE Trans Transp Electrif 8(1):1182–1193. https://doi.org/10.1109/TTE.2021.3093069
24. Modi G, Singh B (2022) Improved cascaded SOGI control for islanding-synchronization in photovoltaic system. IEEE Trans Ind Appl 58(6):6909–6919. https://doi.org/10.1109/TIA.2022.3201808
25. Hsieh C-H, Liang T-J, Chen S-M, Tsai S-W (2016) Design and implementation of a novel multilevel DC–AC inverter. IEEE Trans Ind Appl 52(3):2436–2443. https://doi.org/10.1109/TIA.2016.2527622
26. Mondal B, Karuppaswamy A (2021) An analytical approach to parameter selection for SOGI-Based Fourth-Order Quadrature Signal Generators, In *2021 IEEE 12th energy conversion congress & exposition—Asia (ECCE-Asia)*, Singapore, pp 2363–2368, https://doi.org/10.1109/ECCE-Asia49820.2021.9479349

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/381832242

Rheological Behaviour Analysis of Prepared Nano- Cutting Fluids from Al 2 O 3 and Tio 2

Research · June 2024

CITATIONS
0

READS
25

4 authors:



Anurag Sharma
Delhi Technological University
14 PUBLICATIONS **25** CITATIONS
SEE PROFILE



R. C. Singh
Delhi Technological University
90 PUBLICATIONS **771** CITATIONS
SEE PROFILE



Ranganath M Singari
Delhi Technological University, Formerly Delhi college of Engineering, Delhi, India
197 PUBLICATIONS **1,445** CITATIONS
SEE PROFILE



Dr. Shivaji Bhandarkar
Guru Gobind Singh Indraprastha University
39 PUBLICATIONS **161** CITATIONS
SEE PROFILE

All content following this page was uploaded by Anurag Sharma on 30 June 2024.

The user has requested enhancement of the downloaded file.

STM JOURNALS

Journal of
Polymer & Composites

ISSN: 2321-2810 (Online)
ISSN: 2321-8525 (Print)
Volume 12, Special Issue 1, 2024
DOI (Journal): 10.37591/JoPC

http://engineeringjournals.stmjournals.in/index.php/JoPC/index

Research JoPC

Rheological Behaviour Analysis of Prepared Nano-Cutting Fluids from Al_2O_3 and Tio_2

Anurag Sharma[1],*, R. C. Singh[2], Ranganath M. Singari[2], S. L. Bhandarkar[3]

Abstract

In this research paper, an attempt has been made to characterize and analyse the rheological properties of two samples of nano-cutting fluids. Nanoparticles of Al_2O_3 and TiO_2 were characterized by SEM and EDS. The high peaks showed the respective presence of aluminium, oxygen for alumina and titanium and oxygen for titanium dioxide. The two samples of nano-cutting fluids were prepared. Alpha alumina, α-Al_2O_3 nanoparticles were used for the preparation of the sample of nano-cutting fluid in the ratio of 1% (w/w) with distilled water. TiO_2 nanoparticles were used for the preparation of nano-cutting fluid in the ratio of 1% (w/w) with distilled water. Conventional cutting fluid was made in the ratio of 1:20 (w/w) with distilled water. The three samples were analysed at room temperature 25°C and 65°C. It was found that viscosity of each sample was lower at higher temperature as compared to viscosity at 25°C. The thermal conductivity of each sample was increased at higher temperature by 10.91, 8.42 and 7.4% respectively as compared to low temperature.

Keywords: Nano-cutting fluids, nanoparticles, temperature, viscosity, SEM, TEM

INTRODUCTION

The researchers and scientists are meticulously working in reducing the values of machinability characteristics between the workpiece and cutting tool (i)surface roughness of the machined workpiece, (ii) cutting forces, (iii) coefficient of friction, (iv) tool wear, (v) cutting temperature etc. [1,2]. The attempts are made that coolant used during machining may not generate any effect harmful effects towards the operator and environment. Cryogenic cooling with liquid nitrogen has been used by researchers. This is environment friendly and does not produce any toxic effects on the operator [3, 4]. The conventional cutting fluids might produce fumes, itching to eyes skin etc. and various other types of human health issues problems during machining processes. The direct disposal of used one might not be environmentally friendly. The edible and non-edible oils have been used during machining processes. The health issue problems related to operator and pollution to the environment might have been improved [5-9]. Silver nanoparticles were mixed in conventional cutting fluid and it was found that surface roughness and tool wear reduced [10]. The nano boric powder was mixed in vegetable oil samples of coconut oil and soyabean oil and it was found that nanoparticles reduced the surface roughness and tool wear as compared to dry and wet machining with conventional cutting fluid [11-14]. The vegetable oil and compressed air mixed in a closed chamber are supplied at the interface of the workpiece and cutting tool. The minimum quantity lubrication prevented the excess flow of coolant and supplied the controlled amount of coolant [15]. Nanoparticles are very fine particles of the magnitude of 10^{-9}m. The lubrication, cooling and abrasive properties are dependent on the types of nanoparticles. The heat transfer and tribological

***Author for Correspondence**
Anurag Sharma

[1]Faculty G B Pant DSEU Okhla III, India
[2]Professor Delhi Technological University, India,
[3]Controller of Exams Guru Gobind Singh Indraprastha University, India,

Received Date: January 06, 2024
Accepted Date: March 05, 2024
Published Date: May 04, 2024

Citation: Anurag Sharma, R. C. Singh, Ranganath M. Singari, S. L. Bhandarkar. Rheological behaviour analysis of prepared nano-cutting fluids from Al_2O_3 and TiO2. Journal of Polymer & Composites. 2024; 12(Special Issue 1): S329–S338p.

© STM Journals 2024. All Rights Reserved

properties improved. The faster rate of heat transfer may keep the tool and workpiece at a lower temperature. This further might reduce the coefficient of friction and wear rate [16- 19]. Padmin et al. [20] made samples of nano MoS_2 with (0, 0.25, 0.5, and 0.75%) (w/w) in coconut oil and performed experiments on lathe machine and it was found that 0.5% of composition gave a better results as compared with other compositions. Uysal et al. [21] prepared a composition with MoS_2 1.0% (w/w) with vegetable oil with water emulsion during milling and found better surface finish than conventional cutting fluid. Kumar et al. [22] developed a nanofluid by using multi-walled nano carbon tubes in sunflower by 1% (w/w). The experimentation was performed on a tirbometer and grinder. It was found that tribological properties got improved as compared to a conventional cutting fluid. Amrita et al. [23] prepared various samples of nanofluids by using nanoboric acid, nano molybdenum disulphide with emulsion 0.3% (w/w). It was found that functionalised nano graphite showed good stability in oil emulsion. Sharma et al. [24] prepared different types of samples of nano-cutting fluids with Al_2O_3(0,0.25, 0.5,1.5,2.0 and 3 %) (v/v) with deionised water and found that with 1% composition gave a better performance during turning operations as compared with other compositions. Ramakrishanan et al. [25] found that marginal difference in surface roughness between experimentation performed on lathe machine during 0.5% and 1.0% (w/w) Al_2O_3 nanoparticles with inonized water as coolant. The tool wear was lower with 1.0% (w/w) nano-cuttig fluid as compared to dry and 0.5% (w/w) concentration of Al_2O_3. It was found that very limited literature work is available with the rheological property of nanocuting fluid.The novelty of this research work is that an attempt has been made to focus on characterization of nanoparticles, prepared nano-cutting fluid and the rheological property of nano-cutting fluid like viscosity has been analysed with time period of constant shear rate at two different temperatures like room temperature 25°C and 65°C and compared with conventional cutting fluid. The highly viscous cutting fluids may not be suitable for machining as they may not be able to reach properly at the interface of workpiece and cutting tool and very lean cutting fluid may not provide a good lubricating and cooling effect. The thermal conductivity of nano-cutting fluid was measured and compared with conventional cutting fluid. Therefore, nano-cutting fluids with Al_2O_3 and TiO_2 were prepared characterized and analysed for viscosities at a different temperature

Experimental Methodology

Two different types of nanoparticles like (α-Al_2O_3) alpha-alumina and titanium dioxide (TiO_2) were procured from the manufacturers M/S Plasma ChemGmbH, Berlin, Germany. The nanoparticles are water-soluble. The properties of nanoparticles are given in Table 1. The purity index of nanoparticles was very high in terms of percentage for both the selected nanoparticles.

Table 1. Properties of nanoparticles

S No.	Properties	Nano-Al_2O_3	Nano-TiO_2
1	Average grain size	40nm	8nm
2	Particle size full range	5-100nm	8-25nm
3	Purity	99.99%	99.98%
4	Specific surface	> 10 m^2/g	50±10 m^2/g

The composition was formed by considering the weight by weight ratio with base fluid. In present research work base fluid was distilled water. The empty beaker was weighted on weighing scale. Then scale reading was made zero. The beaker was filled with distilled water and weighed on a scale. Now, nanoparticles were weighed on micro-scale in proportionate to 1% weight of water and mix in beaker.

Nanoparticles of TiO_2 were weighed on micro digital balance with an accuracy of ±0.0001grams. A sample of nano-cutting fluid was made with distilled water in the proportion weight of 1% (weight by weight) of distilled water. Figure. 1 shows a magnetic stirrer for blending of nano-cutting fluid for 60 minutes at 520 rpm at 21°C and rest was given for half an hour. Rest was given for half an hour for checking the sedimentation. The process of magnetic stirring was repeated for 60 minutes. This process was repeated 4-5 times.

© STM Journals 2024. All Rights Reserved



Figure 1. Magnetic stirrer for nano-cutting fluid with nanoparticles TiO_2

Ultrasonification was performed by a water bath ultrasonicator. It was operated with single-phase electric supply at 200-230Voltsat 50Hz as depicted in Figure. 2. One litre distilled water was used for filing the water bath container. 30k Hz utrasonfication waves were generated. Initially, ultrasonification was performed for 30minutes. The rest of five minutes was given. Then it was repeated for five times. Visual inspection did not show any sedimentation. Zeta potential test was conducted. The value was 32 which was in a stable range.



Figure 2. Water bath ultrasonicator

© STM Journals 2024. All Rights Reserved

Nanoparticles of Al_2O_3 were weighed on micro digital balance with an accuracy of ±0.0001grams. A sample of nano-cutting fluid was made with distilled water in the proportion weight of 1% (weight by weight) of distilled water. Magnetic stirring was done for blending of nano-cutting fluid for 60 minutes at 520 rpm at 21°C and rest was given for half an hour. This was repeated 4-5 times and the rest was given overnight for checking any sedimentation. Most of the surfactant used by previous researchers was sodium-based. Tween-20 Biocompatible surfactant has been found, which is non-toxic. Initially, 0.25% was mixed with nano-cutting fluid. Further, it was increased to 1% (weight by weight) of prepared nano-cutting fluid. Magnetic stirring was done at 520rpm at 21°C for 60 minutes. Then ultrasonication was done for 30 minutes and repeated 4-5 times. Zeta potential test was done and value was 21 which was in the stability range.

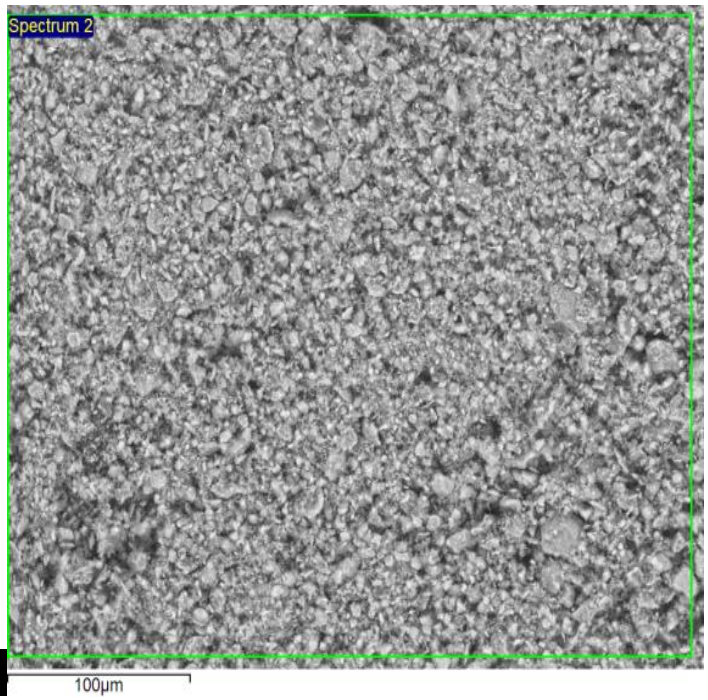
The sample was prepared (w/w) in the ratio of 1:20. 1 part of conventional cutting oil and 20 parts of distilled water. This composition is used during machining processes like turning, milling, grinding, drilling etc.

The thermal conductivity of different cutting fluids was measured by hot wire (KD2 thermal properties analyse, Decagon devices, Inc. USA). The experiment of measuring thermal conductivity of every cutting fluid was repeated five times and an average was calculated for finding out the final value.

Each sample was kept by a dropper at the cup and hob of Rheometer in the enclosed chamber. The temperature was adjusted at room temperature at 25°C and readings were noted for viscosity at the shear rate of 100/second for the time period of 125 seconds. The cup and hob were cleaned by acetone soaked paper. After leaving for drying for some time place another sample of nao-cutting fluid was placed and the chamber was closed and temperature was adjusted to 25°C another sample. Similarly, the samples of nano cutting fluids of Al_2O_3 TiO_2 and conventional cutting fluid were placed and analysed at 65°C.

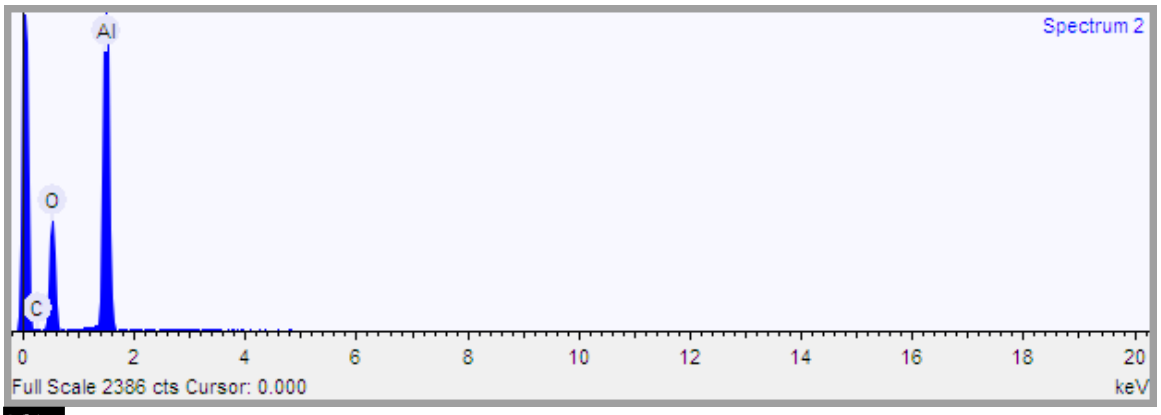
RESULTS AND DISCUSSION

The characterization of nanoparticles Al_2O_3 was done by SEM and EDS images as shown in Figure. 2(a) and (b) at 100µm.Fig. 2 (b) shows high peaks of aluminium and oxygen. It was found that the percentage oxygen was more. The percentage of oxygen and aluminium in weight was depicted as 52.673% and 47.327% respectively.



3(a)

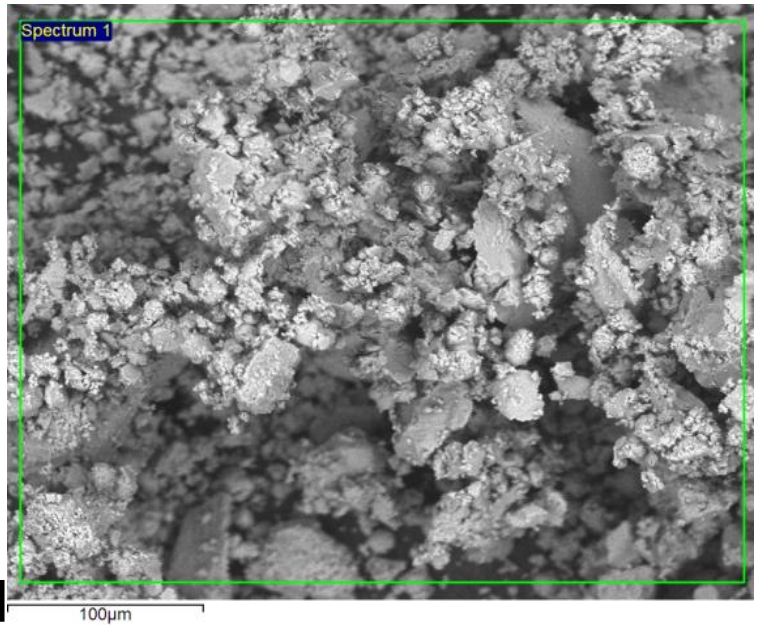
© STM Journals 2024. All Rights Reserved



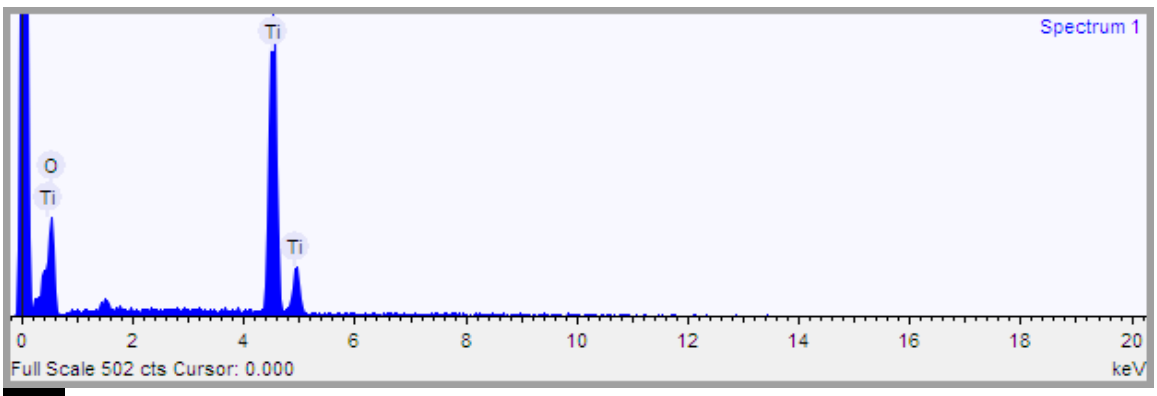
3(b)

Figure 3. (a) SEM image and (b) EDS of nanoparticles Al_2O_3

The characterization of nanoparticles TiO_2 was done by SEM and EDS images as shown in Figure. 4 (a) and (b) respectively with high peaks of titanium and oxygen. It has been shown that percentage of titanium was more. Comparatively, oxygen was progressively lower. The percentage of titanium and oxygen in weight as 49.360% and 50.640%.



4(a)



4(b)

Figure 4. (a) SEM image and (b) EDS of nanoparticles TiO_2

© STM Journals 2024. All Rights Reserved

Figure 5. (a) shows TEM image at magnification of 40000 times and (b) shows TEM image at magnification of 60000 times of the prepared sample of nano-cutting fluid with nanoparticles of Al_2O_3. Some agglomerations of particles were seen.

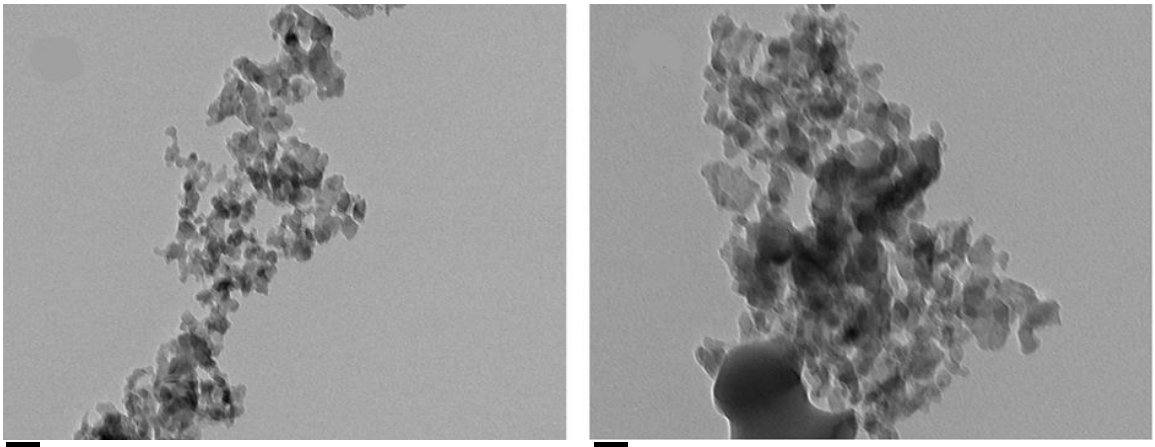


Figure. 5. (a) TEM image at magnification of 40000X 50 nm and (b) 60000X 50nm of prepared Al_2O_3 nano-cutting fluid

Characterization of prepared nano-cutting fluids by performing TEM image of samples Fig. 6 (a) shows TEM image at a magnification of 40000 times and (b) shows TEM image at a magnification of 60000 times of the prepared sample of nano-cutting fluid with nanoparticles of TiO_2. It has been seen that particle size was very fine and no agglomeration of particles was seen.

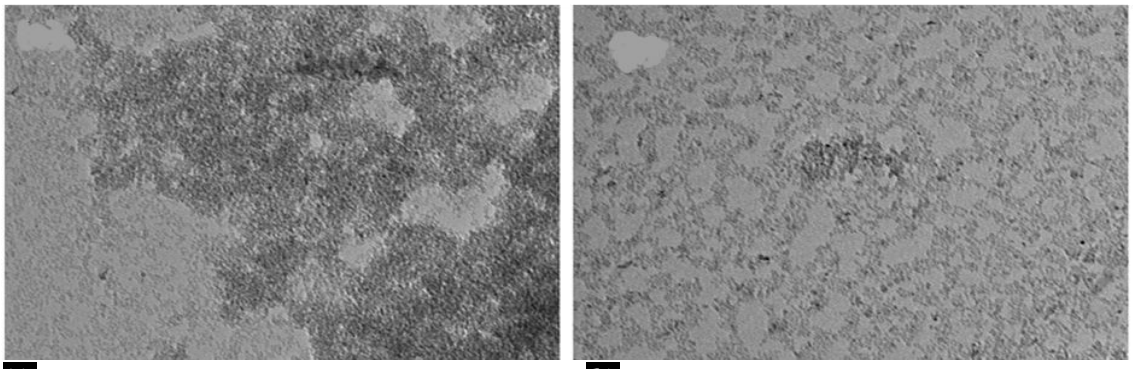


Figure 6. (a) TEM image at magnification of 40000X50nm and (b) 60000X 50nm of prepared TiO_2 nano-cutting fluid

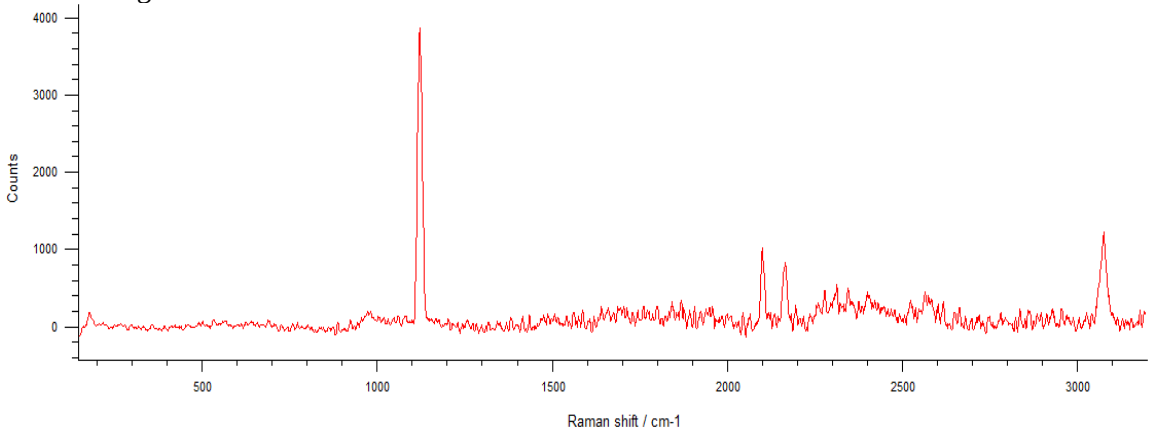


Figure 7. Raman shift pattern for Al_2O_3 in prepared nano-cutting fluid

© STM Journals 2024. All Rights Reserved

It has been shown in Figure. 7 that highest peak of Al_2O_3 was observed at 1100-1200 Raman shift /cm^{-1} for 380counts. This confirmed the presence of Al_2O_3. Another peak was observed at 2200-2300 Raman shift /cm^{-1} for 100counts.

Figure. 8 depicts that the highest peak was observed at 1200-1300 Raman shift /$cm^{-1}$for 980counts. This confirmed the presence of TiO_2 in the prepared nano-cutting fluid. Another high peak was observed at 3200-3300Raman shift /cm^{-1} for 450counts.

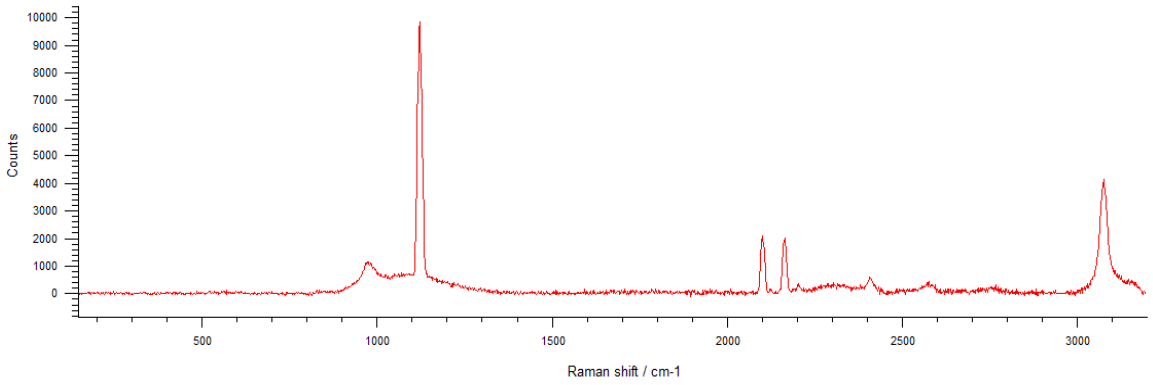


Figure. 8. Raman shift pattern for TiO_2 in prepared nano-cutting fluid

Figure. 8 depicts that the highest peak was observed at 1200-1300 Raman shift /$cm^{-1}$for 980counts. This confirmed the presence of TiO_2 in the prepared nano-cutting fluid. Another high peak was observed at 3200-3300Raman shift /cm^{-1} for 450counts.

It has been depicted in Figure. 9 that at the initial start of rheometer till first 20 seconds the viscosities of AA, TT and CC was higher than respective viscosities at the end of 125 seconds at the shear rate of 100/seconds at 25°C. On comparing the viscosity of nano-cutting fluids AA, TT with CC it was found that viscosity was declined by 91.92 and 90.81% respectively at the viscosity obtained at the end of 125 seconds. In Fig. 10, it was found that viscosities declined at the end of 125 seconds. On comparing the viscosity of nano-cutting fluids AA, TT with CC it was found that viscosity was declined by 94.02 and 86.19% respectively at the velocity obtained the end of 125 seconds at 65°C. On comparing both the Figures. 9 and 10 it was found that respective viscosity of nano-cuting fluids and cutting fluid were lower at higher temperature at 65°C due to weaker intermolecular forces.

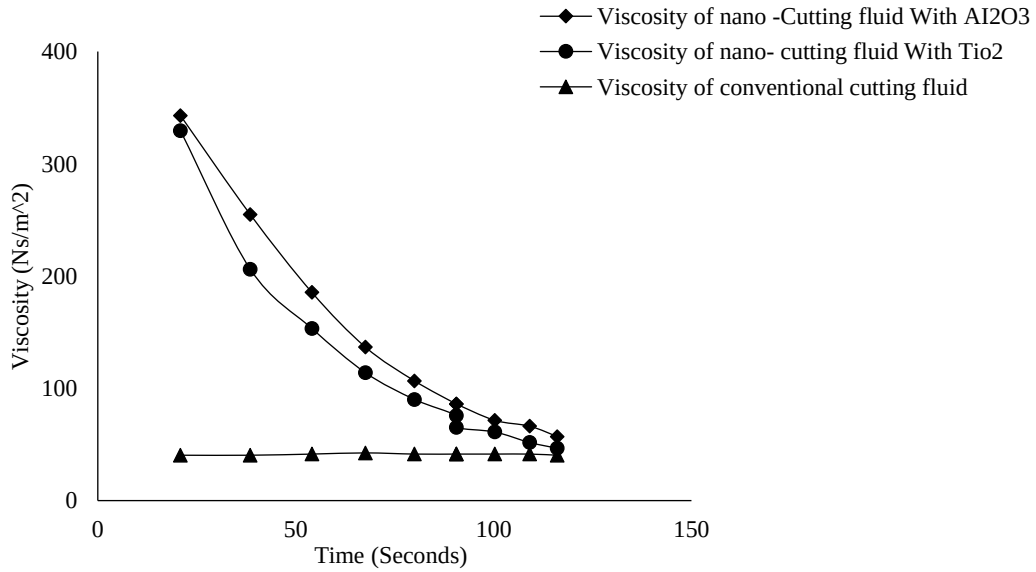


Figure 9. Visocity of cutting fluids with time at room temperature 25°C

© STM Journals 2024. All Rights Reserved

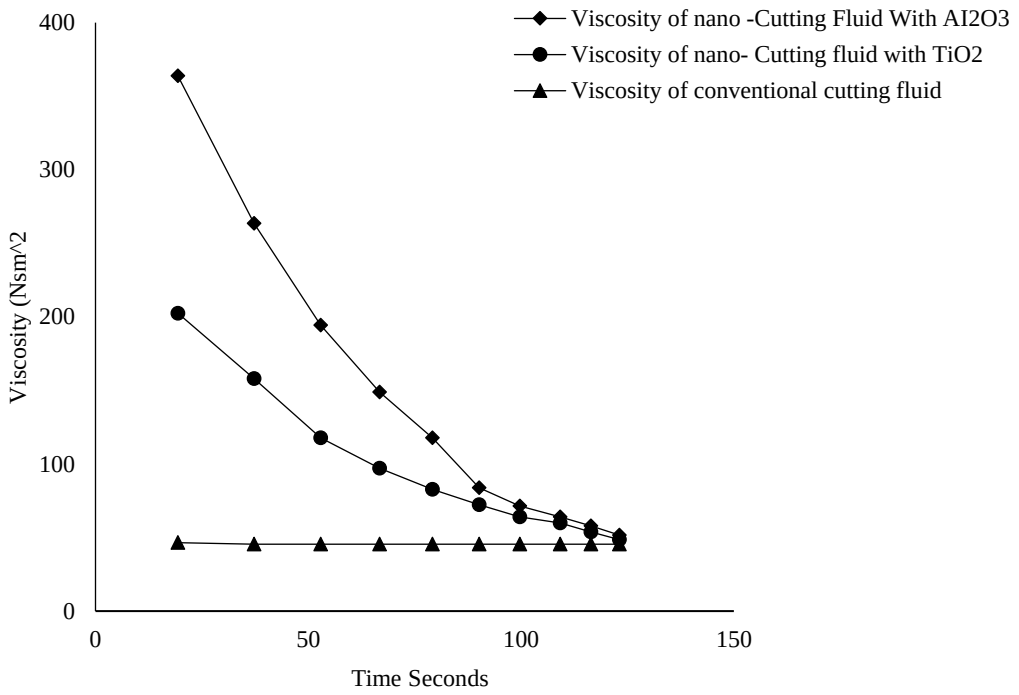


Figure 10. Viscosity of cutting fluids with time at 65°C

The thermal conductivity of cutting fluids was found to be more at higher temperature of 65°C on comparing with temperature of 25°C as depicted in Figure. 10. The thermal conductivity was enhanced by 7.42, 10.91 and 8.42 % respectively for CC, AA and TT at higher temperature of 65°C. The inclination in thermal conductivity was higher for nano-cutting fluid with Al_2O_3 due to the property of material.

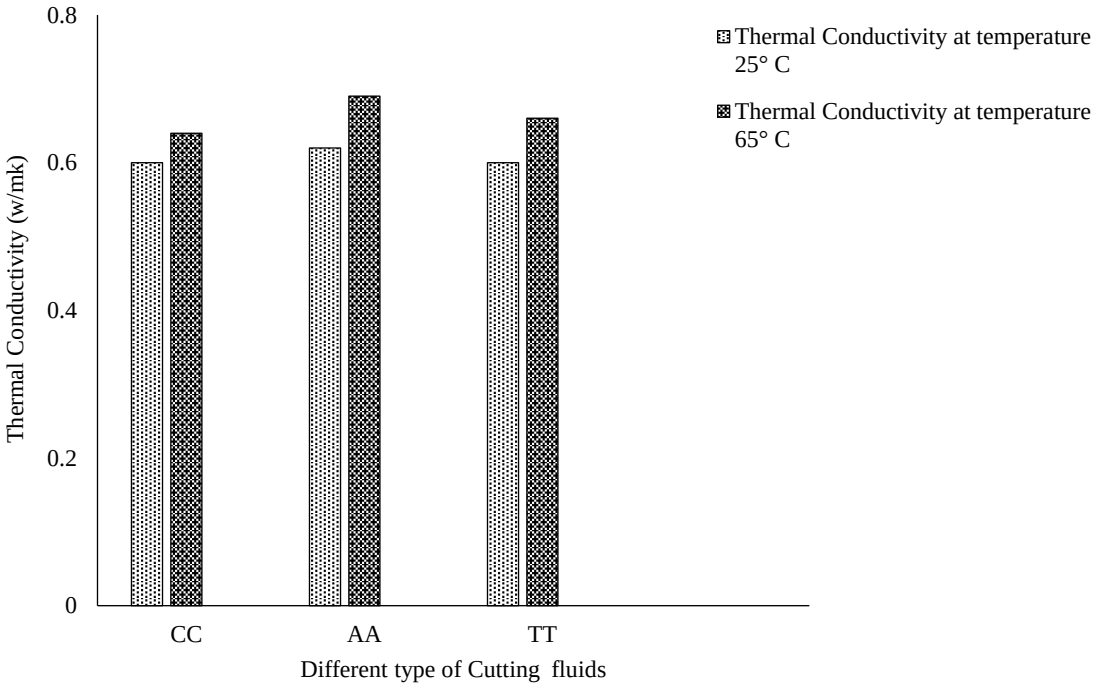


Figure 10. Thermal conductivity of nano-cutting fluids and conventional cutting fluids

© STM Journals 2024. All Rights Reserved

CONCLUSIONS

The nano-cutting fluids were prepared from two distinct nanoparticles and after characterization the samples were testes for change in viscosities on a dedicated rheometer. The following conclusions were summarized.

1. Biocompatible surfactant was used during the preparation of nano-cutting fluid with Al_2O_3.
2. The temperature influenced the viscosity of prepared nano-cutting fluids.
3. The viscosity of nano-cutting fluids and conventional cutting fluids declined with the increasing time period at shear rate of 100/second.
4. The viscosity of nano-cutting fluids and cutting fluids declined with increasing temperature.
5. The nano-cutting fluid with Al_2O_3 depicted more stability in viscosity change.
6. The thermal conductivities increased during higher temperature.
7. The thermal conductivity of nano-cutting fluid with Al_2O_3 was higher than other cutting fluids.

Abbreviations

w/w: Weight by weight
AA: Nano-cutting fluid with Al_2O_3
TT: Nano-cutting fluid with TiO_2
CC: Conventional cutting fluid (1:20)

Acknowledgements

Authors are highly thankful to workshop and laboratories facilities shared by Delhi Technological University and Indian Institute of Technology, Delhi (India)

REFERENCES

1. Byrne G, Scholta E (1993) Environmentally clean machining processes - a strategic approach. CRIP Annals- Manufacturing Technology 42(1):471-474
2. Hong S Y, Zaho (1999) Thermal aspects, material considerations and cooling strategies in cryogenic machining. Clean Production Processes 1:107-116
3. Agrawal R, Kumar A, Singh S, Sharma K. (2022) Recent advances and future perspectives of lignin biopolymers. Journal of Polymer Research, 29(6).
4. AlSaidi R A, Alamri H R, Sharma K, Al-Muntaser A A (2022). Insight into electronic structure and optical properties of ZnTPP thin films for energy conversion applications: experimental and computational study. Materials Today Communications, 32, 103874.
5. Lawal S A, Choudhary I A, Nukman Y (2012) Application of vegetable oil-based metal working fluids in machining ferrous metals- A Review. International Journal of Machine Tools & Manufacture 52:1-12
6. Singh P K, Singh P K, Sharma, K (2022). Electrochemical synthesis and characterization of thermally reduced graphene oxide: Influence of thermal annealing on microstructural features. Materials Today Communications, 32, 103950.
7. Shrikant U G, Patil N G (2018) Experimental investigations into turning of hardened AISI 4340 steel using vegetable based cutting fluids under minimum quantity lubrication. Procedia Manufacturing 20:18-23
8. Sharma A, Chaturvedi R, Sharma K, Saraswat M. (2022). Force evaluation and machining parameter optimization in milling of aluminium burr composite based on response surface method. Advances in Materials and Processing Technologies, 8(4), 4073-4094.
9. Nizammuddin M, Agrawal S M, Patil N (2018) The effect of kanranja based soluble cutting fluid on chips formation in orthogonal cutting process of AISI 1045 steel. Procedia Manufactturing 20: 12-17
10. Chaturvedi R, Sharma A, Sharma K, Saraswat M. (2022). Tribological behaviour of multi-walled carbon nanotubes reinforced AA 7075 nano-composites. Advances in Materials and Processing Technologies, 8(4), 4743-4755.

© STM Journals 2024. All Rights Reserved

11. Pavani P N L, Pola R, Srikiran S (2015) Performance evaluation and optimization of nanoboric acid powder weight percentage mixed with vegetable oil using Taguchi approach. Journal of Mechanical Science and Technology 29: 4877-4883
12. Krishna P V, Srikant R R, Rao D N (2010) Experimental investigation on the performance of nano boric suspensions in SAE 40 and coconut oil during turning of AISI 1040 steel. International Journal of Machine Tools & Manufacture 50: 911-916
13. Chetan, Behera B C, Ghosh S, Rao P V (2016) Application of nanofluids during minimum quantity lubrication: A case study in turning process.Tribology International 101:234–246
14. Shokrani Y, Shckarian E (2015) Application of nanofluids in machining processes- A Review. Journal Nanoscience Technology 2: 59-63
15. Liu Z, Zhanguo, S. U, Abed M, Chaturvedi R, Feyzbaxsh M, Salavat A K.(2022). A comparative thermodynamic and exergoeconomic scrutiny of four geothermal systems with various configurations of TEG and HDH unit implementations. Applied Thermal Engineering, 216, 119094.
16. Zhang M, Wang X, Liu W (2009) Performance and antiwear mechanism of Cu nanoparticles as lubricating oil additives. Industrial Lubrication Tribology 61(6): 311-318.
17. Chan C Y, W. B. Lee and H. Wang (2013) Enhancement of surface finish using water – miscible nano cutting fluid in ultra precision turning. International Journal of Machine Tools & Manufacture73:62-70
18. Sun Y, Gong L, Li B (2016) Performance evaluation of nanofluid MQL with vegetable – based oil and ester oil as base fluids in turning. International Journal Advanced Manufacturing Technology 83 ((9-12): 2083-2089
19. Hai T, Alshahri A H, Mohammed A S, Sharma, Almujibah H R, Metwally A ,Ullah, M. (2023). Performance assessment and multiobjective optimization of a biomass waste-fired gasification combined cycle for emission reduction. Chemosphere, 334, 138980.
20. Uysal A, Demiren F, Atlan E (2015) Applying minimum quantity lubrication (MQL) method on milling of martensitic stainless steel by using Nano MoS_2 reinforced vegetable cutting fluid. Procedia social and Behavioral Sciences, 195: 2742-2747.
21. Kumar K M, Ghosh A (2016) Assessment of cooling lubrication and wettablity characteristics of nano engineered sunflower oil as cutting fluid and its impact on SQCL grinding performance Journal of Material Processing Technology, 237: 55-64.
22. Amrita M, Shariq S A, Manoj Gopal C (2014) Experimental investigation on application of emufsifier oil based nano cutting fluid in metal cutting process. Procedia Engineering, 97: 1115-1124
23. Hai T, Ali M A, Dhahad H A, Alizadeh A A, Sharma A, Almojil S F, Wang D (2023) Optimal design and transient simulation next to environmental consideration of net-zero energy buildings with green hydrogen production and energy storage system. Fuel, 336, 127126.
24. Sharma A K , Singh R K , Dixit A R, Tiwari A K. 2016. Characterization and experimental investigation of Al_2O_3 nanoparticle based cutting fluid in turning AISI 1040 steel under minimum quantity lubrication (MQL). Materials Today, Proceedings 3(22):1899-1906
25. Ramakrishnan H, Balasundaram R, Selvaganapathy P , Santhakumari, Sivasankaran P, Vignesh P (2019) Experimental investigation of turning Al 7075 using Al_2O_3 nano-cutting fluid: ANOVA and TOPSIS approach. S/N applied sciences 1: 1639 https://doi.org/10.1007/s42452-019-1664-0.

© STM Journals 2024. All Rights Reserved

View publication stats

Check for updates

Received: 14 February 2024 | Revised: 20 June 2024 | Accepted: 24 June 2024

DOI: 10.1002/cta.4175

ORIGINAL ARTICLE

WILEY

Second generation current conveyor based capacitorless floating memristor emulator

Navnit Kumar[1] | **Manjeet Kumar**[1] | **Neeta Pandey**[1] | **Shahram Minaei**[2]

[1]Department of Electronics and Communication Engineering, Delhi Technological University (DTU), Delhi, India

[2]Department of Electrical and Electronics Engineering, Dogus University, Istanbul, Turkey

Correspondence
Neeta Pandey, Department of Electronics and Communication Engineering, Delhi Technological University (DTU), Delhi, 110042, India.
Email: n66pandey@rediffmail.com

Abstract

This article offers a novel flux controlled floating memristor emulator based on second-generation current conveyor (CCII). The floating memristor is designed using two CCIIs, one resistor, and one PMOS transistor. The presented memristor emulator does not need external capacitance. The proposed designed circuit exhibits pinch hysteresis loops in voltage–current plane up to 2 GHz frequency. The performance of the circuit under consideration is evaluated using 180 nm CMOS technology parameter by the SPICE simulator. The circuit requires a DC power supply of ±1.2 V and exhibits a power consumption of 0.766 mW. Furthermore, the resilience of the planned circuit is assessed by examining process corner fluctuation, supply voltage variation, temperature variation, and transistor size variations. In addition, a Schmitt trigger circuit and high order filters based on designed memristor are used to confirm the operation of the proposed design at high and low frequency, respectively.

KEYWORDS

memristor, non-volatile nature, pinch hysteresis loop, Schmitt trigger circuit, second-generation current conveyor (CCII)

1 | INTRODUCTION

Leon O. Chua introduces memristor as the fourth circuit element after the capacitor, inductor, and resistor in 1971.[1] The memristor is recognized as a nonlinear resistor with memory, serving as a crucial component that establishes a connection between flux and charge.[2] The memristor has desirable characteristics such as configurable resistance and memory functionality. Additionally, memristor has the ability to retain their prior state even in the absence of power, thereby exhibiting a non-volatile property.[2] A memristor may be used in cellular networks,[3] sensors,[4] chaotic circuit,[5,6] signal processing,[7–9] digital logic,[10] and neuromorphic systems.[11] Recently, various remarkable efforts have been undertaken to develop resilient memristor models for commercial applications, including spintronic memristors,[12] organic memristors,[13] carbon nanotube,[14] and amorphous silicon memristors.[15] However, these models are constrained by limits pertaining to cost, manufacturing, and performance. Hence, researchers put their effort to design a memristor emulator using active building block (ABB). The memristor developed by ABB has a comparatively low cost and easy to fabricate.

ABB-based memristor emulators have been described in a variety of research papers published in recent years.[16–43] Memristor circuits developed in previous studies[16,17] are based on programmable analog circuits. It consists of a digital potentiometer, a microprocessor, and an analog-to-digital converter (ADC). Despite this, performance is restricted due

 © 2024 John Wiley & Sons Ltd.

to the limiting resolution of the digital potentiometer. Fitch et al.[18] presented a memristor emulator circuit that was based on the light-dependent resistor (LDR). This circuit is made up of three operational amplifiers, 11 resistors, one LDR, one capacitor, and one diode. The performance of the memristor was hindered by LDR characteristics while operating at higher frequencies. By integrating four current feedback operational amplifiers (CFOA), two diodes, four resistors, and four capacitors, Abuelama'atti and Khalifa[19] were able to develop a memristor that was easy to use and had a wide range of applications. A voltage-controlled resistor (VCR) constructed with six resistors, two capacitors, two operation amplifiers (op-amps), and an NMOS transistor, together with a second-generation current conveyor (CCII), and an external capacitor are components that make up the memristor circuit that was developed by Hussein and Fouda.[20] An off-the-shelf floating memristor was proposed by Sozen and Cam,[21] and it is comprised of four CCIIs, three OTAs, one capacitor, and six resistors. It is necessary to use an additional voltage source in order to bias the current of one of the OTAs. Valsa et al.[22] designed a memristor employing five Op-amps, 14 resistors, one JFET, and one capacitor. To function properly, many memristor emulators[23–32] need additional circuits, like a multiplier circuit, in addition to both active and passive components. It is possible that the adoption of additional circuits will increase the complexity of the memristor circuit while simultaneously lowering the frequency at which it operates. The floating memristor described in this article[33] does not have any multiplier circuits; hence, it just has a single grounded capacitor and two OTAs. In spite of this, it is operated up to 400 kHz frequency. The voltage inverting metamutator (VIM) was developed by Arundeepakvel et al.[34] One VIM, one diode, two resistors, one capacitor, and one inductor make up its constituent parts. According to the literature, circuits that are based on passive inductors are both cumbersome and costly. The fact that the memristor emulators described in literature[35,36] make use of a variety of bias voltages makes it essential to have a layout of a high quality for their production.[33] To put this into perspective, the floating memristor emulator described in Vista and Ranjan[37] only requires one voltage difference transconductance amplifier (VDTA), one capacitor, and one resistor in order to operate up to 50 MHz. Further, Kumar et al. developed floating memristor[38] utilizing one dual-x current conveyor transconductance amplifier (DXCCTA), four resistors, and one grounded capacitor. It is able to operate up to 60 MHz. In addition, the floating memristor that is discussed in Ranjan et al.[39] has a maximum operating frequency of 10 MHz. There is one grounded capacitor, one current conveyor transconductance amplifier (CCTA), and three resistors utilized in the design of the emulator. A current-differencing transconductance amplifier (CDTA) and OTA were used in the construction of a floating memristor, which was described in Yadav et al.[40] It is possible to operate at a frequency of up to 1 MHz in incremental and decremental mode by making just minor adjustments to the design. A grounded memristor emulator may be found in the research literature.[41–43] The findings of the study indicate that grounded memristors are not desirable for use in complex circuits.[37,39]

It is observed from above literature that most of the reported memristor circuit employ parallel plate capacitor and additional circuit that leads to reduce in maximum operating frequency. There are a limited number of memristor emulators available that do not use analog multipliers and passive resistors. These emulators often include a hysteresis loop that is restricted to frequencies within the kilohertz range. Few memristors are designed for grounded mode of operation, which are not suitable for complex applications. This article aims to develop a capacitorless floating memristor emulator using two CCIIs, one resistor, and one PMOS transistor. The maximum operating frequency of the proposed memristor is 2 GHz. In addition, the proposed circuit is free from addition circuits like analog multiplier circuits, passive capacitor, diodes, LDR, passive inductor, and ADC, which is desirable in terms of the integration viewpoint. Further, memristor-based high-order high-pass filter circuit, memristor-based high-order low-pass filter circuit, and memristor-based Schmitt trigger circuit are implemented by the proposed memristor emulator. The subsequent sections of the article are structured in the following manner: Section 2 presents the circuit description of the proposed circuit. Simulation results of designed memristor emulators are presented in Section 3. Comparative study of the designed memristors with existing ABB based memristors is discussed in Section 4. Section 5 reports applications of the designed circuits. Finally, Section 6 contains the paper's conclusion.

2 | CIRCUIT DESCRIPTION

This section presents a basic overview of CCII. Following that, the proposed capacitorless floating memristor circuit is explained.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.1 | Second generation current conveyor (CCII)

Second generation current conveyor (CCII) is a basic current-mode ABB. It provides wide dynamic range, high linearity, and high frequency performance as compared to performances of voltage mode ABB counterparts. Circuit symbol and CMOS structure of CCII are given in Figure 1A,B, respectively. Figure 1 shows that CCII has three ports, namely, X, Y, and Z. Port relations of the CCII are given as follows.

$$I_X = I_Z; V_X = V_Y; I_Y = 0 \tag{1}$$

It is observed from Equation (1) that the voltage at Y port is conveyed to X port while current fed into X port is conveyed to Z port.

2.2 | Proposed floating memristor emulator

Schematic of designed floating memristor is demonstrated in Figure 2. It consists of two CCIIs, one resistor and one PMOS transistor. The designed circuit utilizes parasitic capacitance (C_g) associated at gate terminal of M_P. CCIIs, resistor, and parasitic capacitance are the components of the feedback network. The memristor current passes through the transistor (M_P) is controlled by the feedback network.

In practice, CCIIs consist parasitic components at input and output terminals. Hence, the proposed floating memristor with parasitic components is depicted in Figure 3.

Figure 3 shows the parasitic components. R_{z1} and C_g are parasitic resistance at Z_1 terminal of CCII and capacitance associated at gate terminal of transistor M_P, respectively. The capacitance C_g is the equivalent capacitor of parasitic capacitances at the gate terminal of PMOS including C_{z1}, C_{gdP}, C_{gsP}, and C_{gbP}. Now, the maximum capacitance at gate terminal of M_P may be expressed as

$$C_g = C_{z1} + W_P L_P C_{OX} + C_{gsov} + C_{gdov} \tag{2}$$

where W_P, L_P, Cox, C_{gsov}, and C_{gdov} are channel width, channel length, oxide capacitance (per unit area), gate-source overlap capacitance, and gate-drain overlap capacitance, respectively. Now, using Sedra et al.,[44] we can derive the values for C_{gdP} and C_{gsP} as

$$C_{gsov} = C_{gdov} = W_P L_{OV} C_{OX} \tag{3}$$

Now, substitute Equation (3) into Equation (2), we obtain following expression of C_g as

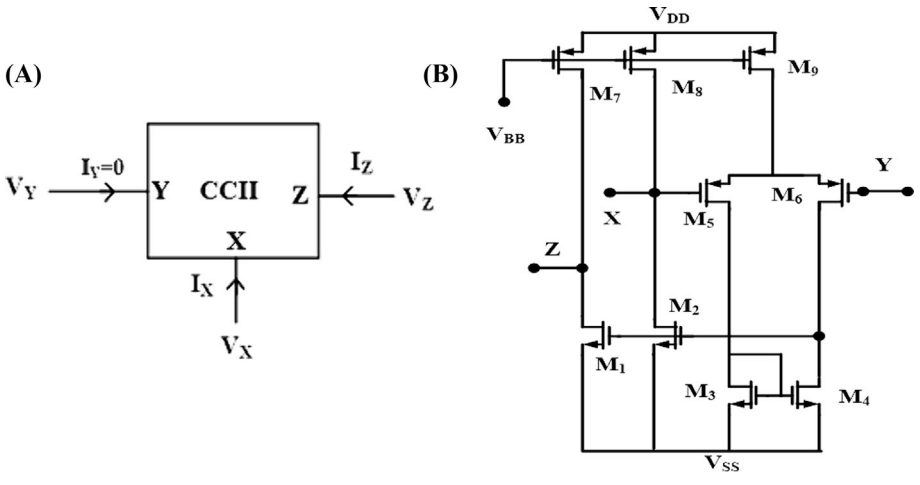


FIGURE 1 CCII circuit. (A) Symbol and (B) CMOS structure.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

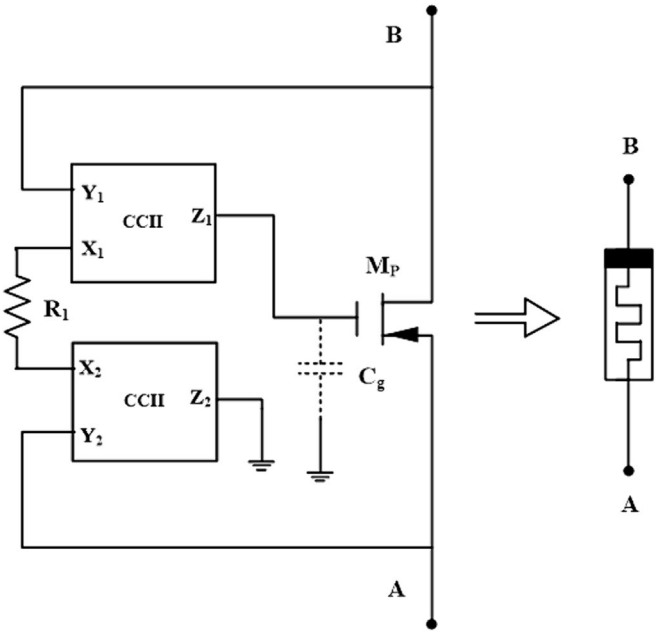


FIGURE 2 Proposed floating memristor circuit.

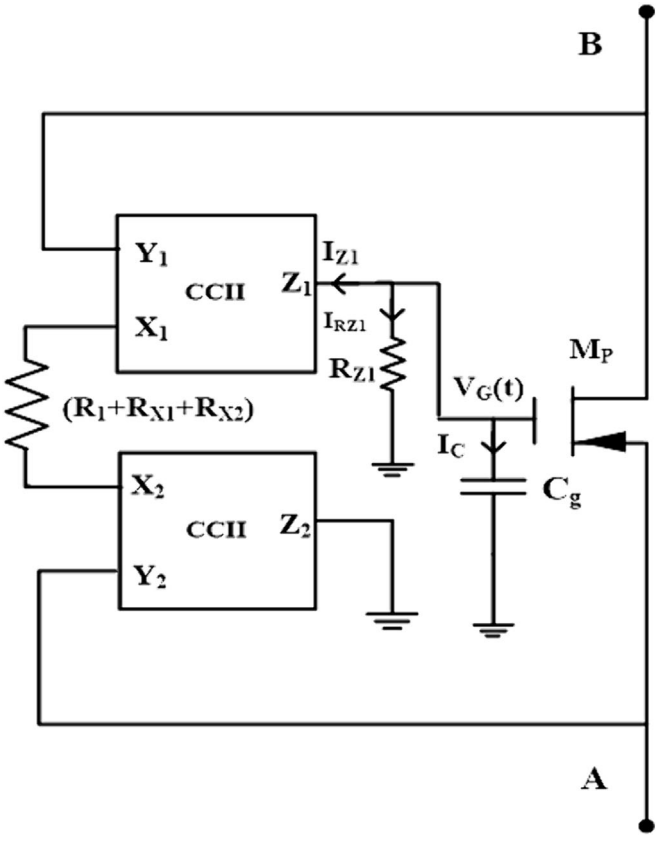


FIGURE 3 Proposed memristor circuit with parasitic components.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$C_g \cong C_{z1} + W_P L_P C_{OX} + 2W_P L_{OV} C_{OX} \tag{4}$$

Substitute the value of $C_{OX} = \frac{\varepsilon_{ox}}{t_{ox}}$ in Equation (4); thus, expression of C_g may be expressed as

$$C_g = C_{z1} + W_P L_P \left(\frac{\varepsilon_{ox}}{t_{ox}}\right) + 2W_P L_{OV} \frac{\varepsilon_{ox}}{t_{ox}} \tag{5}$$

where t_{ox} and ε_{ox} are thickness of the oxide layer, and permittivity of the silicon oxide, respectively. Now, expression of $I_C(t)$ may be derived from from Figure 3 as

$$I_C(t) = -(I_{Z1} + I_{RZ1}) \tag{6}$$

The magnitude of current flows through I_{Z1} and I_{ZR1} of the proposed circuit may be expressed as follow.

$$I_{Z1} = \frac{V_{in}(t)}{(R_1 + R_{X1} + R_{X2})} \tag{7}$$

$$I_{ZR1} = \frac{V_G(t)}{R_{Z1}} \tag{8}$$

Substitute the values of I_{Z1} and I_{ZR1} in Equation (6), we get following expression of $I_C(t)$ as

$$I_C(t) = -\frac{V_{in}(t)}{(R_1 + R_{X1} + R_{X2)}} - \frac{V_G(t)}{R_{Z1}} \tag{9}$$

Substitute $I_C(t) = C_g \frac{dV_G(t)}{dt}$ in Equation (9), we get following expression as

$$\frac{dV_G(t)}{dt} = -\frac{V_{in}(t)}{C_g(R_1 + R_{X1} + R_{X2)}} - \frac{V_G(t)}{C_g R_{Z1}} \tag{10}$$

Rearrange the first-order differential equation as given in Equation (10) as

$$\frac{dV_G(t)}{dt} + \frac{V_G(t)}{C_g R_{Z1}} = -\frac{V_{in}(t)}{C_g(R_1 + R_{X1} + R_{X2)}} \tag{11}$$

Now, integrating factor (IF) may be obtained from Equation (11) as

$$IF = e^{\int \frac{1}{C_g R_{Z1}} dt} = e^{\left(\frac{t}{C_g R_{Z1}}\right)} \tag{12}$$

Now, solution of Equation (11) may be expressed as

$$V_G(t) e^{\left(\frac{t}{C_g R_{Z1}}\right)} = \left\{ -\frac{1}{C_g(R_1 + R_{X1} + R_{X2})} \int V_{in}(t) e^{\left(\frac{t}{C_g R_{Z1}}\right)} dt \right\} + c \tag{13}$$

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

where c is integration constant. Let $V_{in}(t) = V_m Sin\omega t$ where V_m represents the amplitude of the input voltage and ω represents the angular frequency; then, flux is $\varphi(t) = \int V_{in}(t)dt = -\frac{V_m Cos\omega t}{\omega}$, and $V_G(t)$ may be written as

$$V_G(t)e^{\left(\frac{t}{C_g R_{Z1}}\right)} = \left\{-\frac{V_m}{C_g(R_1 + R_{X1} + R_{X2})}\int Sin\omega t e^{\left(\frac{t}{C_g R_{Z1}}\right)}dt\right\} + c \tag{14}$$

Now, let $I = \int Sin\omega t e^{\left(\frac{t}{C_g R_{Z1}}\right)}dt$, then I can be found as

$$I = \int Sin\omega t e^{\left(\frac{t}{C_g R_{Z1}}\right)}dt = \frac{e^{\frac{t}{C_g R_{Z1}}}\left(C_g R_{Z1} Sin\omega t - \omega\left(C_g R_{Z1}\right)^2 Cos\omega t\right)}{1 + \omega^2\left(C_g R_{Z1}\right)^2} \tag{15}$$

Now, substitute the value of I in Equation (14), we get following expression of $V_G(t)$ as

$$V_G(t) = -\frac{V_m}{C_g(R_1 + R_{X1} + R_{X2})}\left\{\frac{C_g R_{Z1} Sin\omega t - \omega\left(C_g R_{Z1}\right)^2 Cos\omega t}{1 + \omega^2\left(C_g R_{Z1}\right)^2}\right\} + ce^{-\left(\frac{t}{C_g R_{Z1}}\right)} \tag{16}$$

If input angular frequency $\omega \gg \frac{1}{C_g R_{Z1}}$, then expression of $V_G(t)$ may be rewritten as

$$V_G(t) = \frac{V_m Cos\omega t}{\omega C_g(R_1 + R_{X1} + R_{X2})} + ce^{-\left(\frac{t}{C_g R_{Z1}}\right)} \tag{17}$$

Now, $\frac{V_m Cos\omega t}{\omega}$ is known as $-\varphi(t)$; thus, $V_G(t)$ may be expressed in terms of $\varphi(t)$ as

$$V_G(t) = -\frac{\varphi(t)}{C_g(R_1 + R_{X1} + R_{X2})} + ce^{-\left(\frac{t}{C_g R_{Z1}}\right)} \tag{18}$$

Since the value of $\frac{1}{C_g R_{Z1}}$ is high, $e^{-\left(\frac{t}{C_g R_{Z1}}\right)}$ term decays quickly towards zero; thus, Equation (18) may be rewritten as

$$V_G(t) \cong -\frac{\varphi(t)}{C_g(R_1 + R_{X1} + R_{X2})} \tag{19}$$

Since transistor M_P is used in the linear region, thus, current $i_D(t)$ of transistor M_P is written as follows:

$$i_D(t) = \mu_P C_{OX}\left(\frac{W}{L}\right)_P\left\{\left(v_{SGP} - \left|V_{Tp}\right|\right)v_{SDP} - \frac{V_{SDP}^2}{2}\right\} \tag{20}$$

By substituting $V_G(t) = -\frac{\varphi(t)}{C_g(R_1 + R_{X1} + R_{X2})}$, $V_A(t) = \frac{v_{in}(t)}{2}$, and $V_B(t) = -\frac{v_{in}(t)}{2}$ in Equation (20), expression for drain current may be obtained as follows:

$$i_D(t) = \mu_P C_{OX}\left(\frac{W}{L}\right)_P\left\{\left(\frac{V_{in}(t)}{2} + \frac{\varphi(t)}{C_g(R_1 + R_{X1} + R_{X2})} + V_{Tp}\right)V_{in}(t) - \frac{V_{in}^2(t)}{2}\right\} \tag{21}$$

The expression for the drain current, denoted as $i_D(t)$, of the constructed memristor emulator for small inputs $V_{in}(t)$ may be expressed in a simpler form as

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$i_D(t) = \mu_P C_{OX}\left(\frac{W}{L}\right)_P\left(\frac{\varphi(t)}{C_g(R_1+R_{X1}+R_{X2})}+V_{Tp}\right)V_{in}(t) \tag{22}$$

Thus, we may create the memductance $(w(\varphi(t)))$ of the designed circuit as

$$w(\varphi(t)) = \frac{i_D(t)}{V_{in}(t)} = \mu_P C_{OX}\left(\frac{W}{L}\right)_P\left(\frac{\varphi(t)}{C_g(R_1+R_{X1}+R_{X2})}+V_{Tp}\right) \tag{23}$$

More investigation into the frequency analysis of the suggested memristor emulator may be conducted recalling that $\varphi(t) = -\frac{V_m \text{Cos}\omega t}{\omega}$ and substituting into Equation (23) the memductance $w(\varphi(t))$ may be derived as

$$w(\varphi(t)) = \underbrace{\left\{\mu_P C_{OX}\left(\frac{W}{L}\right)_P V_{Tp}\right\}}_{\text{Linear time invarient part}} + \underbrace{\left\{\mu_P C_{OX}\left(\frac{W}{L}\right)_P \frac{V_m}{(R_1+R_{X1}+R_{X2})C_g\omega}\text{Cos}(\omega t-\pi)\right\}}_{\text{Linear time varient Part}} \tag{24}$$

Now, we may derive the ratio of amplitudes of linear time-variant part to linear time-invariant part as

$$\alpha = \frac{V_m}{V_{Tp}(R_1+R_{X1}+R_{X2})C_g\omega} \tag{25}$$

where α of the developed memristor depends upon amplitude of the input signal, threshold voltage of M_P, passive resistance, parasitic resistances, input angular frequency, and parasitic capacitance. From (25), it is obvious that for high frequencies of operation, α is decreased, and memductance value is dominated by the linear time-invariant part; thus, the hysteresis curve is converted to a simple straight line acting similar to a linear resistor.

3 | SIMULATION RESULTS AND DISCUSSIONS

A comprehensive examination of the designed capacitorless memristor is conducted by SPICE simulator, using 180 nm technology. The supply voltage is selected to be ±1.2 V. The aspect ratio of transistor M_1, M_2, M_3, M_4, M_5, M_6, M_7, M_8, M_9, and M_P is $\frac{2.4\,\mu m}{0.18\,\mu m}$, $\frac{2.4\,\mu m}{0.18\,\mu m}$, $\frac{0.65\,\mu m}{0.18\,\mu m}$, $\frac{0.65\,\mu m}{0.18\,\mu m}$, $\frac{0.35\,\mu m}{0.18\,\mu m}$, $\frac{0.35\,\mu m}{0.18\,\mu m}$, $\frac{2.4\,\mu m}{0.18\,\mu m}$, $\frac{2.4\,\mu m}{0.18\,\mu m}$, $\frac{2.4\,\mu m}{0.18\,\mu m}$, and $\frac{50\,\mu m}{0.18\,\mu m}$, respectively. The bias voltage (V_{BB}) is set at a value of 0.42 V. Current and voltage ratio of CCII are depicted in Figure 4. In order to design the memristor, passive component is selected as $R_1 = 10\ k\Omega$. Now, a device is considered to be a memristor if it possesses the following three fingerprints.[45] (1) "When driven by a bipolar periodic signal the device must exhibit a pinch

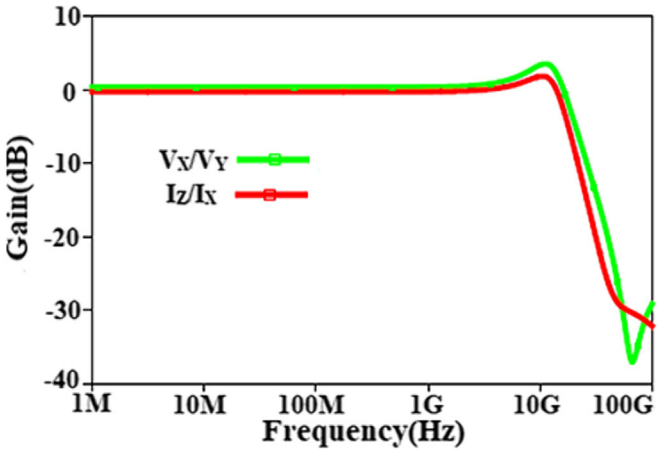


FIGURE 4 Frequency response of CCII.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

hysteresis loop in the voltage-current plane, assuming the response is periodic." (2) "Starting from critical frequency, hysteresis loop area should decrease monotonically as the exciting frequency increases." (3) "The pinch hysteresis loop should shrink to a single valued function when frequency tends to infinity." First fingerprint of the offered memristor is displayed in Figure 5A–H. It illustrates the progressive transformation of the pinch hysteresis loop (PHL) of the memristor under consideration, spanning the frequency range of 25 MHz to 2 GHz. Figure 5A–H is also depicting that the current passing across the memristor tends to approach zero as the input voltage is reduced to 0. The second and third fingerprints of the designed memristor are displayed in Figure 5E–H. Figure 5E–H shows that when the frequency of the applied signal increases from 200 MHz, there is a reduction in the area of the lobes seen in the PHL curve.

Given that the memristor is classified as a resistor, it can be inferred that the phase difference between the input current and input voltage is zero. The transient response of the memristor emulator under consideration, operating at a frequency of 200 MHz, is shown in Figure 6. The observation demonstrates that the magnitude of the input current is zero when the differential input voltage reaches zero.

The memristor circuit under consideration exhibits temperature-dependent variations in its current during practical implementation. Hence, the impact of temperature fluctuations ranging from −45°C to +90°C on the PHL of the

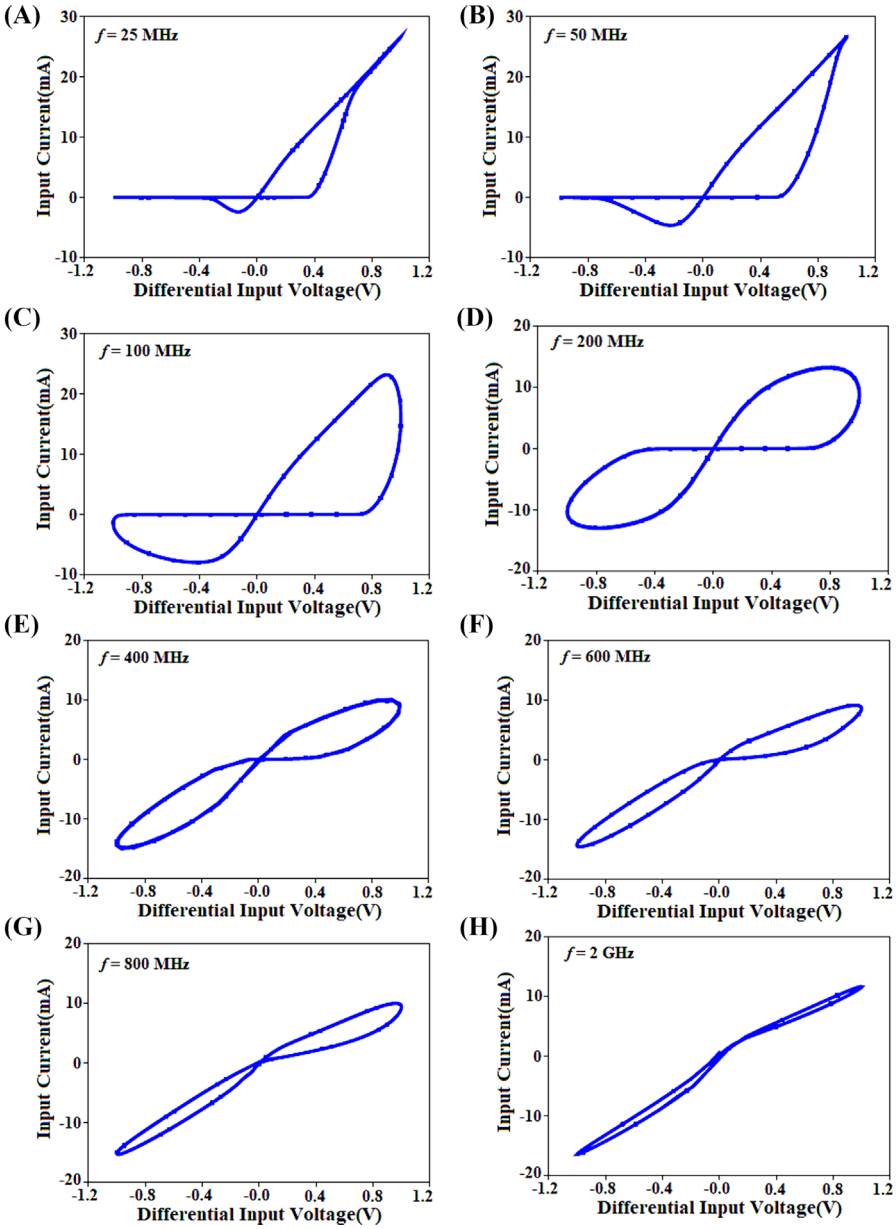


FIGURE 5 PHL for the frequency range 25 MHz to 2GHz.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

memristor at a frequency of 200 MHz is shown in Figure 7. The observed phenomenon indicates that there is a drop in memristor current as temperatures rise, which may be attributed to a loss in carrier mobility. Further, Figure 8 shows the PHL of the proposed memristor at frequency of 200 MHz for various aspect ratios of the transistor M_P. It shows that the amount of current going through the memristor change from 6.15 to 25.96 mA as the aspect ratio of the transistor M_P goes up (30 μm/0.18 μm to 90 μm/0.18 μm).

The performance of the presented memristor emulator may be varied by the fluctuations in the bias voltage (V_{BB}) of CCII. Hence, Figure 9 shows the V-I characteristics of the developed memristor when the V_{BB} shifts from +0.35 to +0.45 V. Figure 9 is also depicting that increasing in the bias voltage from +0.35 to +0.45 V results change in input current between 13.3 and 15.4 mA.

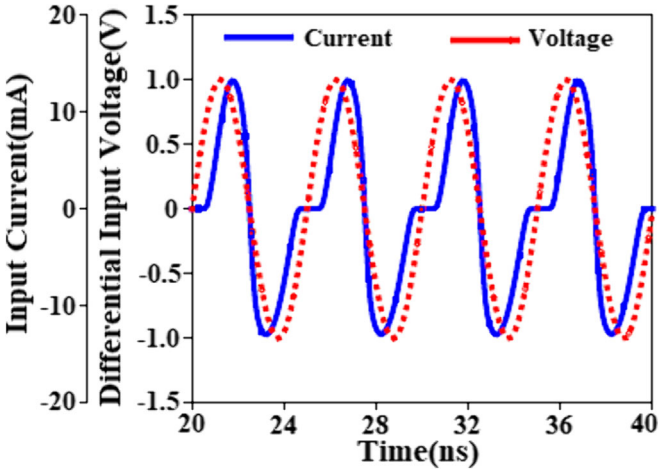


FIGURE 6 Transient response of the memristor at 200 MHz frequency.

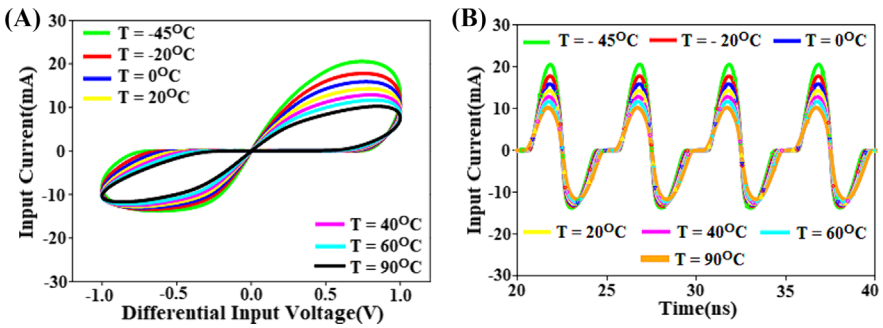


FIGURE 7 Temperature variation in (A) PHL and (B) time domain response.

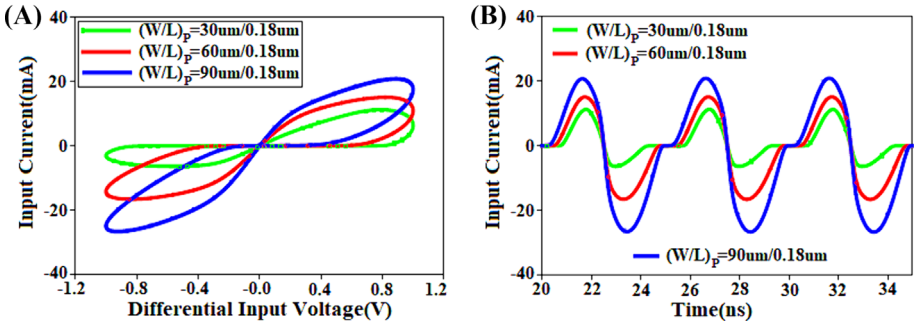


FIGURE 8 Transistor (M_P) size variation in (A) PHL and (B) time domain response.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The efficient operation of the planned memristor might be hampered due to fluctuations in the DC power supply. So, Figure 10 shows the suggested memristor's V-I characteristics curve when the supply voltage shifts from ±1.1 to ±1.3 V. Figure 10 displays that increasing in the supply voltage from ±1.1 to ±1.3 V results deviations in input current between 11.1 and 15.06 mA.

In monolithic integration, process corner has a major impact on how well the proposed circuit works. It has been meticulously examined in the slow slow (SS), fast fast (FF), and nominal nominal (NN) process corners. The PHL at different process corner of the presented circuit are depicted in Figure 11. The process, voltage, and temperature (PVT) variations are achieved by changing the supply voltage and temperature between (±1.1 to ±1.3 V) and (−45°C to

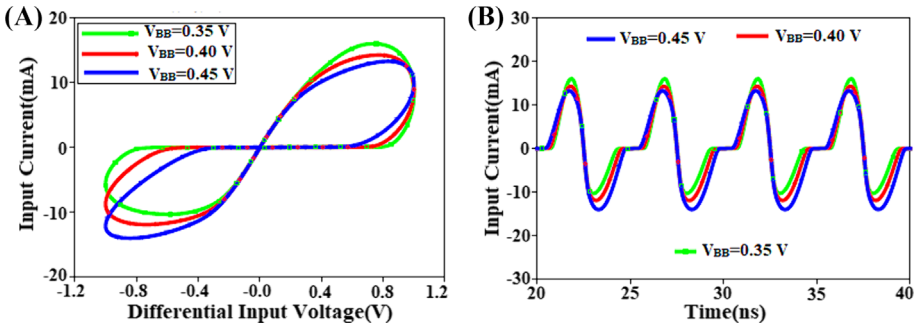


FIGURE 9 Variation of bias voltage of CCII in (A) PHL and (B) time domain response.

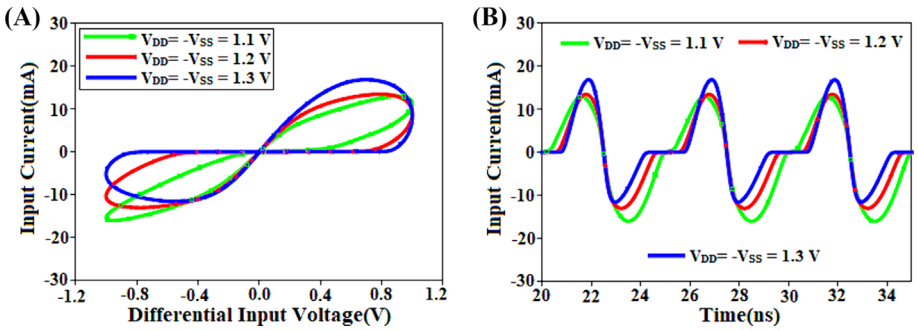


FIGURE 10 Supply voltage fluctuation in (A) PHL and (B) time domain response.

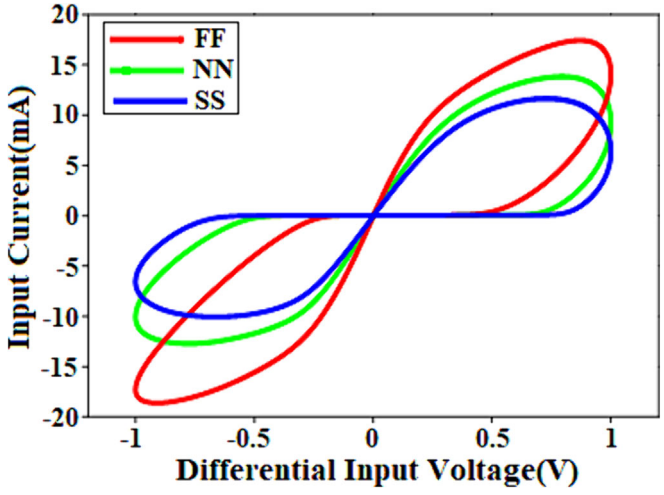


FIGURE 11 PVT analysis of the memristor.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

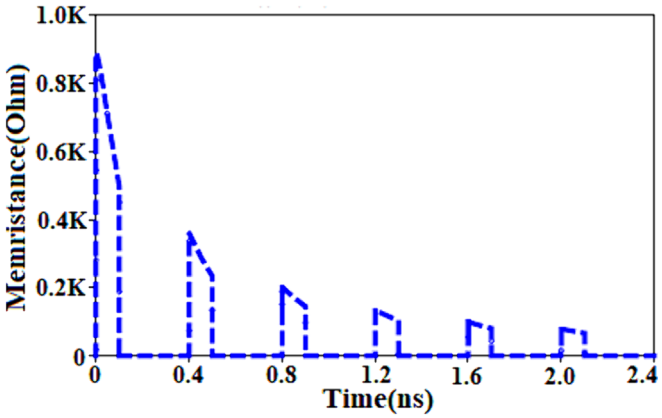


FIGURE 12 Memristance of the memristor.

+90°C) respectively. Figure 11 shows that the current through a memristor in SS mode is less than the current through the same device in FF mode. The proposed memristor exhibits a constricted PHL in each process corner, despite the fact that the hysteresis loop may exhibit some small-area fluctuation.

In addition to the PHL, the non-volatility of memristance is a distinguishing characteristic of the memristor emulator. The memristor has the potential to retain its programmed memristance value for a prolonged period even if no input signal is present. Figure 12 displays the memristance of the designed memristor circuit. This response is obtained by providing a pulse signal with amplitude of 1 V, having a pulse width of 100 ps, and time period of 400 ps has been applied at the input of the floating memristor. Figure 12 illustrates that a small amount of variation of memristance in ON period of pulse due to non-idealities associated with CCII.

4 | COMPARISON

Table 1 illustrates a comparative study between the performances of the designed memristor emulator and an existing memristor emulator. The comparison is conducted based on many factors, including the number of active blocks, the number of passive components, the technology node, power dissipation, supply voltage, need of passive capacitor, the mode of operation, and the maximum operating frequency.

There are following conclusion may be drawn from Table 1 as follows:

a. The memristor emulator reported in previous studies[18–43] utilizes parallel plate capacitor. In opposite, the proposed memristor uses parasitic capacitor.
b. The memristor emulator described in previous works[18,19,23–32,34] includes addition circuit such as diodes, LDR, inductor, and multiplier circuits, which may reduce the maximum operating frequency. However, the designed memristor emulator is free from additional circuits.
c. Power consumption in realization of memristor reported in literature[38,39,42,43] are 4, 7.5, 2.6, and 9.56 mW, respectively. However, the designed topology of memristor consumes 0.766 mW power.
d. The memristor emulator suggested in previous works[18,22,23,25–28,30–32,34,35,40–43] realize for grounded mode of operation. In opposite, the proposed circuit may be used in floating mode of operation.
e. Excessive active (more than 2) and passive components are utilized in the topology of memristor emulators presented in literature.[18–24,30,31] In contrast, the proposed memristor circuit employs two CCIIs and one floating resistor.
f. The memristor emulator presented in previous studies[21,26,28,29,32,35,36,38–41,43] requires OTA circuit. OTA restricts the operating frequency of the designed memristor at upper frequency range. However, the proposed circuit does not utilize OTA.
g. Maximum operating frequency achieved in literature[18–36,40,41] are within 10 MHz. In distinction, maximum operating frequency of the designed floating memristor circuit is 2 GHz.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 The proposed memristor emulator is compared to an existing ABB-based memristor emulator.

Reference	No. of active block	Passive elements R and C	Parameter (nm)	Power supply (V)	Floating/ grounded	Power dissipation	Additional circuits	Need of passive capacitor	Max. operating freq.
Fitch et al.[18]	Op-amp-3	11 and 1	-	±15	Grounded	NA	LDR-1, Diode-1	Yes	280 Hz
Abuelama'atti et al.[19]	CFOA-4	4 and 4	-	NA	Floating	NA	Diode-2	Yes	6 kHz
Hussein and Fouda[20]	CCII-1, Op-amp-2, NMOS-1	6 and 3	130	±1.5	Floating	NA	-	Yes	1 MHz
Sozen and Cam[21]	CCII-4, OTA-3	6 and 1	-	±15	Floating	NA	-	Yes	10 kHz
Valsa et al.[22]	Op-amp-5, JFET-1	14 and 1	-	±12	Grounded	NA	-	Yes	2 Hz
Lopez et al.[23]	AD844AN-2	2 and 1	-	±10	Grounded	NA	Multiplier-1	Yes	160 kHz
Lopez et al.[24]	AD844AN-3	5 and 1	-	±10	Floating	NA	Multiplier-2	Yes	20.2 kHz
Yesil et al.[25]	DDCC-1	2 and 1	350	±1.5	Grounded	NA	Multiplier-1	Yes	1 MHz
Ayten et al.[26]	CBTA-1	2 and 1	180	±0.9	Grounded	NA	Multiplier-1	Yes	460 kHz
Sánchez-López and Aguila-Cuapio[27]	AD844AN-1	2 and 1	-	NA	Grounded	NA	Multiplier-1	Yes	860 kHz
Babacan et al.[28]	MO-OTA-1	1 and 1	180	±1.25	Grounded	NA	Multiplier-1	Yes	1000 Hz
Petrovic[29]	VDTA-1	2 and 1	180	±0.9	Floating	NA	Multiplier-1	Yes	2 MHz
Xie et al.[30]	Op-amp-5	10 and 1	-	NA	Grounded	NA	Multiplier-2	Yes	100 Hz
Wei et al.[31]	Op-amp-6	14 and 1	-	NA	Grounded	NA	Multiplier-1	Yes	15 kHz
Pal et al.[32]	VDTA-2	2 and 1	45	±0.9	Grounded	3.956 mW	Multiplier-1	Yes	53 kHz
Kanyal et al.[33]	OTA-2	0 and 1	180	±1.2	Floating	NA	-	Yes	400 kHz
Arundeepakvel et al.[34]	VIM-1	2 and 1	180	±1.65	Grounded	NA	Diode-1, Inductor-1	Yes	10 Hz
Bhardwaj and Srivastava[35]	VDCC-1 OTA-1	2 and 1	180	NA	Grounded	NA	-	Yes	1 MHz
Fatih and Babacan.[36]	OTA-1, MOSFET-3	0 and 2	180	±1.25	Floating	NA	-	Yes	24 Hz
Kumar et al.[38]	DXCCTA-1	4 and 1	90	±1.2	Floating	4 mW	-	Yes	60 MHz

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Reference	No. of active block	Passive elements R and C	Parameter (nm)	Power supply (V)	Floating/ grounded	Power dissipation	Additional circuits	Need of passive capacitor	Max. operating freq.
Ranjan et al.[39]	CCTA-1	3 and 1	250	±1.5, −1	Floating	-	-	Yes	10 MHz
Yadav et al.[40]	CDBA-1 OTA-1	0 and 1	180	±0.9	Floating	NA	-	Yes	1 MHz
Ranjan et al.[41]	DVCCTA-1	3 and 1	250	±1.25	Grounded	NA	-	Yes	1 MHz
Kumar et al.[42]	CCII-1, MOSFET-1	1 and 1	-	±1.5	Grounded	2.6 mW	-	Yes	40 MHz
Raj et al.[43]	CCII-1, OTA-1	1 and 1	180	±1.2	Grounded	9.567 mW	-	Yes	26.3 MHz
Proposed	CCII-2, MOSFET-1	1 and 0	180	±1.2	Floating	0.766 mW	-	No	2 GHz

Abbreviation: NA, not available.

10970007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5 | APPLICATIONS

The operation of the memristor emulators suggested in this study has been evaluated via several applications: (1) Schmitt trigger circuit using flux controlled memristor, (2) low pass filter using designed memristor, and (3) high pass filter using designed memristor.

5.1 | Schmitt trigger circuit using proposed flux controlled memristor

Schmitt trigger circuit is usually used to get rid of noise in digital circuits as part of signal shaping. It is also used for conversion of analog input signal to digital signal. Figure 13 illustrates a memristor-based Schmitt trigger circuit, serving as an exemplification of the intended memristor. The circuit configuration comprises a single operational amplifier, suggested memristor, and a floating resistor. Now, the higher and lower threshold voltages of a Schmitt trigger may be mathematically represented as:

$$V_{TH} = +V_{Sat}\frac{R_M}{R_2} \tag{26}$$

$$V_{LH} = -V_{Sat}\frac{R_M}{R_2} \tag{27}$$

where R_M is memristance of the developed memristor circuit. The dependence of the higher threshold voltage and lower threshold voltage of a memristor-based Schmitt trigger circuit on the supply voltage of the operational amplifier, memristance (R_M), and the feedback resistor (R_2) may be noticed from the preceding equations. The supply voltage for op-amp is chosen as ±15 V. The resistance values of R_1 and R_2 are selected as 10 kΩ. Amplitude and frequency of input sinusoidal voltage source are considered as 1 V and 200 MHz, respectively. Output and input voltage waveforms of the designed Schmitt Trigger circuit are presented in Figure 14. It may be noticed from Figure 14 that a sinusoidal signal is converted into to square wave signal by the designed circuit. In addition, hysteresis curve of the designed Schmitt trigger circuit is illustrated in Figure 15. Figure 15 illustrates transfer characteristics of the Schmitt trigger circuit. It is observed from Figure 15 that the upper threshold voltages are 367.40, 384.425, and 421.849 mV. Moreover, the lower threshold voltages are obtained as −183.68, −235.92, and −277.698 mV.

5.2 | Low pass filter using designed memristor

A low pass filter (LPF) using designed memristor is illustrated in Figure 16. The frequency response of the LPF is depicted in Figure 17. It is produced by applying a sinusoidal voltage source of 1 V with capacitance value $C_1 = C_2 = C_3 = C_4 = C = 1.85$ pF, and $R_{M1} = R_{M2} = R_{M3} = R_{M4} = R_M$ (using $R_1 = 200$ kΩ resulting in memristance value of approximately 230 MΩ at low frequencies). The transfer function of the given LPF may be expressed as

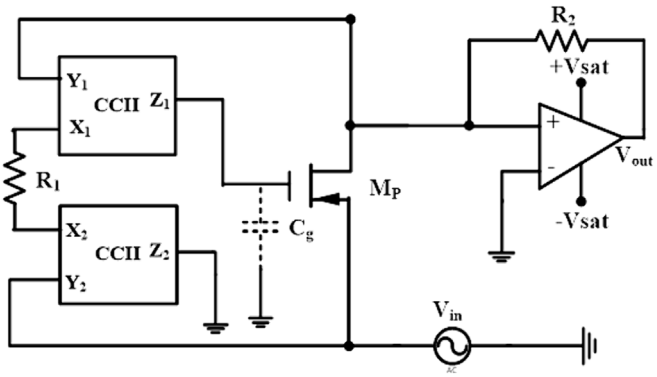


FIGURE 13 Schmitt trigger circuit using designed memristor.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

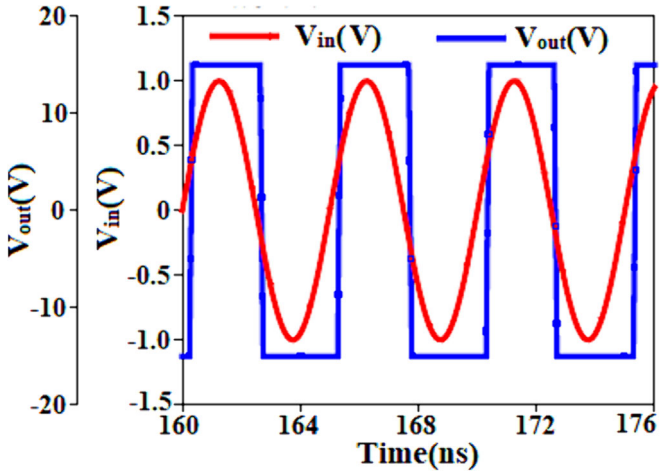


FIGURE 14 Input and output voltage waveform of Schmitt trigger circuit.

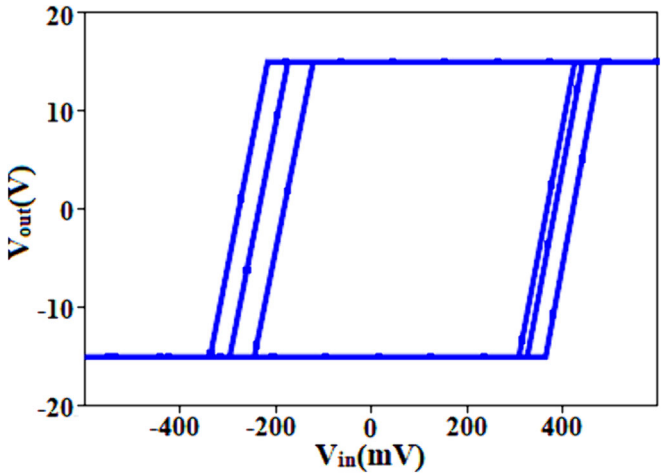


FIGURE 15 Hysteresis curve of Schmitt trigger circuit.

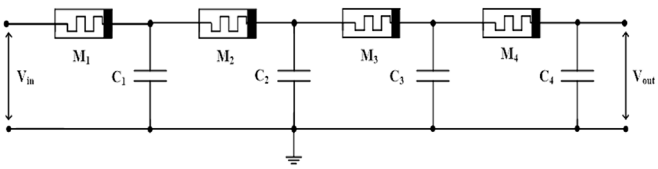


FIGURE 16 LPF using proposed memristor.

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{s^4R_M^4C^4 + 7s^3R_M^3C^3 + 15s^2R_M^2C^2 + 10sR_MC + 1} \tag{28}$$

It may be noticed from Figure 17 that the cut off frequency and roll-off rate are 22.12 Hz and −80 dB/decade, respectively. The result of the designed LPF exhibits the expected behavior that validates the applicability in biomedical fields. Monte Carlo analysis of frequency response of memristor-based low pass filter for 400 runs with 5% tolerance in passive components is shown in Figure 18. Figure 19 displays output noise of the memristor based LPF at 100 mHz and 1 MHz are −106.86 and −169.38 dB, respectively. Total power consumption of the designed filter is 3.06 mW.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

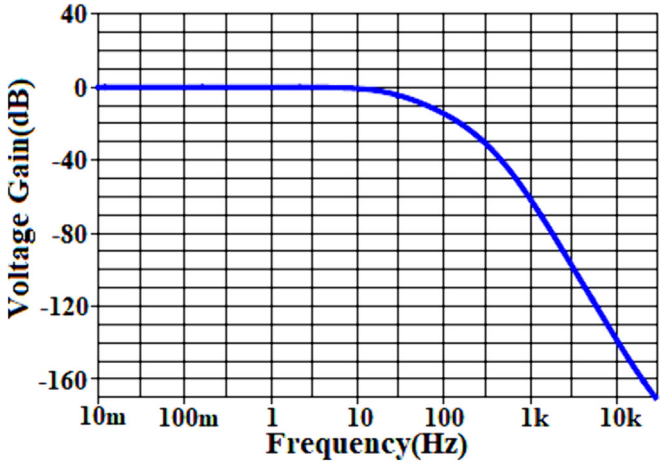


FIGURE 17 Frequency response of LPF.

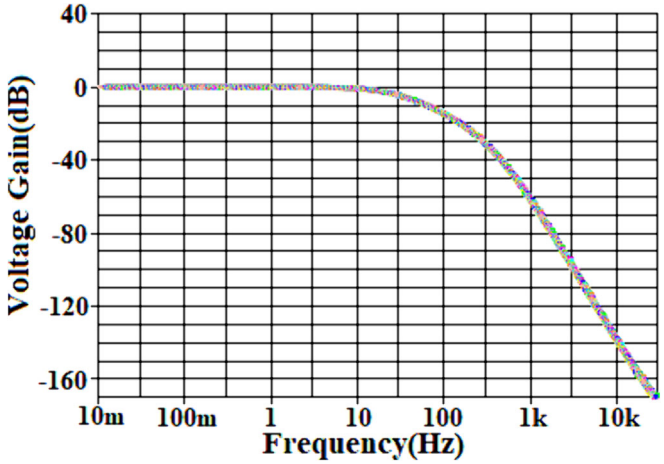


FIGURE 18 Monte Carlo analysis of frequency response of LPF.

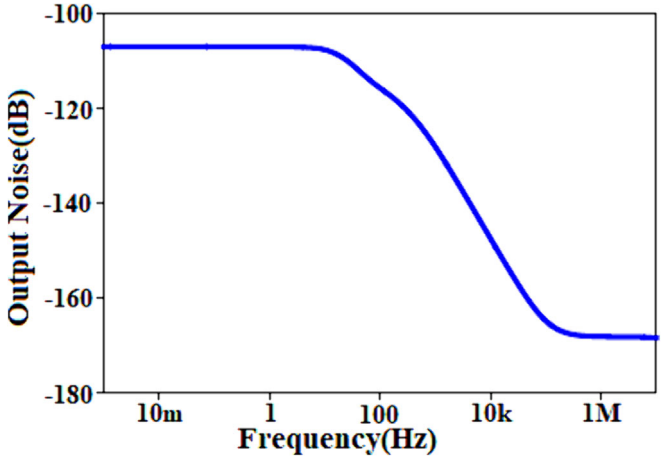


FIGURE 19 Output noise of LPF.

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

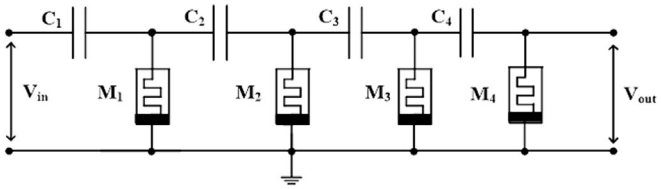


FIGURE 20 HPF using proposed memristor.

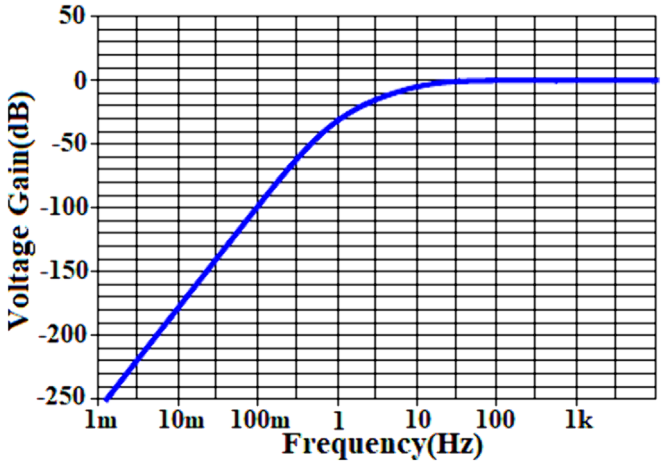


FIGURE 21 Frequency response of HPF.

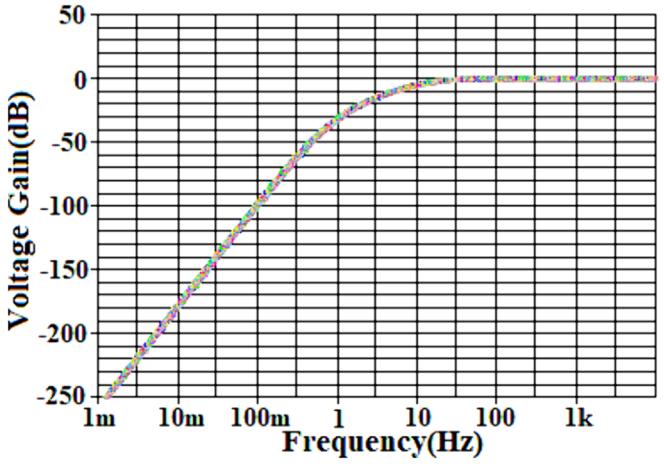


FIGURE 22 Monte Carlo analysis of frequency response of HPF.

5.3 | High pass filter using proposed memristor

The proposed floating memristor may be used as grounded memristor by providing $V_B(t) = 0$ V. A voltage mode high pass filter (HPF) using designed memristor is illustrated in Figure 20. The functionality of the HPF is tested using component values: $C_1 = C_2 = C_3 = C_4 = 200$ pF, and $R_{M1} = R_{M2} = R_{M3} = R_{M4} = R_M$ (using $R_1 = 200\,k\Omega$). The transfer function of the memristor based HPF may be derived as

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

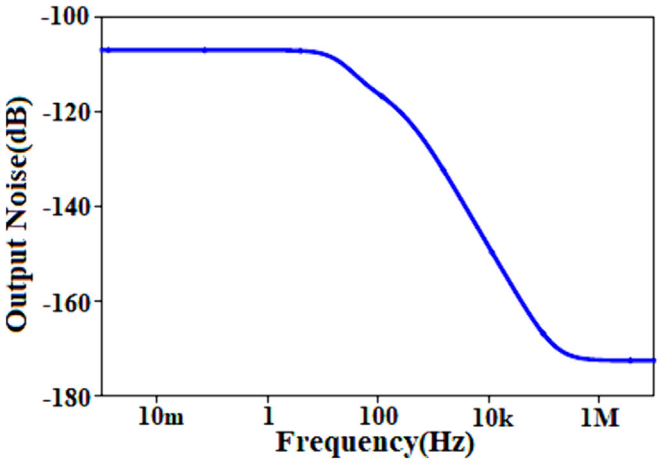


FIGURE 23 Output noise of HPF.

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{s^4 R_M^4 C^4}{s^4 R_M^4 C^4 + 10 s^3 R_M^3 C^3 + 15 s^2 R_M^2 C^2 + 7 s R_M C + 1} \tag{29}$$

Frequency response of HPF is illustrated in Figure 21. It may be noticed from Figure 21 that the cut off frequency and roll-off rate of the filter are 14.29 Hz and 80 dB/decade respectively. Monte Carlo analysis of frequency response of the HPF for 400 runs with 5% tolerance in passive components is shown in Figure 22. Figure 23 depicts output noise of the memristor based HPF at 100 mHz and 1 MHz are −112.89 and −191.077 dB, respectively. Total power consumption of the designed HPF is 3.06 mW.

6 | CONCLUSION

A capacitorless floating memristor emulator is designed in this manuscript. The reported emulator utilizes two CCIIs, one resistor and one PMOS transistor. Implementation of the proposed circuits does not include complex circuits like analog multiplier circuits, passive inductors, diodes, and ADC circuits. The developed memristor exhibits satisfying operation for changing the wide range of temperature. It shows a constricted PHL in each process corner, even though the pinch hysteresis loop may display some small-area fluctuation. Further, the PHL of the proposed memristor for various aspect ratios of the transistor (M_P) is also demonstrated. Since designed emulator excludes external capacitor, it exhibits PHL in V-I plane up to 2 GHz frequency. A Schmitt trigger circuit is constructed utilizing the suggested memristor to validate the workings of the created memristor at high frequency. It may be useful to get rid of noise in digital circuits at higher frequencies. In addition, high-order LPF and HPF circuits are also designed to illustrate the performances of the designed circuit at low frequencies. Results of the designed filters shows the expected behavior, which validates the applicability in biomedical fields. Thus, the proposed circuit exhibits the following advantages. (a) It is simple in structure. (b) It utilizes one passive component. (c) It is capacitorless. (d) The maximum operating frequency of the designed memristor emulator is 2 GHz. (e) The power consumption of the proposed memristor emulator is 0.766 mW. (f) It is suitable for complex applications. (g) It is free from additional circuit such as passive inductor, diodes, ADC, and analog multiplier. (h) It is suitable for monolithic integration. (i) It shows the applicability in biomedical fields.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Manjeet Kumar https://orcid.org/0000-0001-6578-9741

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

REFERENCES

1. Chua LO. Memristor: the missing circuit element. *IEEE Trans Circuit Theory*. 1971;18(5):507-512. doi:10.1109/TCT.1971.1083337
2. Kim H, Sah MP, Yang C, Cho S, Chua LO. Memristor emulator for memristor circuit applications. *IEEE Trans Circuits Syst I: Reg Pap*. 2012;59(10):2422-2431.
3. Duan S, Hu X, Dong Z, Wang L, Mazumder P. Memristor-based cellular nonlinear/neural network: design, analysis, and applications. *IEEE Trans Neural Netw Learn Syst*. 2015;26(6):1202-1213. doi:10.1109/TNNLS.2014.2334701
4. Adeyemo A, Mathew J, Jabir A. Efficient sensing approaches for high-density memristor sensor array. *J Comput Electron*. 2018;17(3):1285-1296. doi:10.1007/s10825-018-1176-y
5. Jie Z, Guo Y, Xu L, Zhu X, Yang J. Hyperchaotic circuit design based on memristor and its application in image encryption. *Microelectron Eng*. 2022;265:111872. doi:10.1016/j.mee.2022.111872
6. Luo J, Chen X. Transmission synchronization of multiple memristor chaotic circuits via single input controller and its application in secure communication. *Integr J*. 2023;90:40-50. doi:10.1016/j.vlsi.2023.01.005
7. Lu Y, Liang Q, Huang X. Parameters self-tuning PID controller circuit with memristors. *Int J Circ Theory Appl*. 2018;46(1):138-154. doi:10.1002/cta.2316
8. Yu D, Iu HHC. A floating memristor emulator based relaxation oscillator. *IEEE Trans Circuits Syst I: Reg Pap*. 2014;61(10):2888-2896. doi:10.1109/TCSI.2014.2333687
9. Fouda ME, Radwan AG. Memristor-based voltage-controlled relaxation oscillators. *Int J Circuit Theory Appl*. 2014;42(10):1092-1102. doi:10.1002/cta.1907
10. Li Z, Yu D, Ye Z, Iu HH, Fernando T. Memristor-based logic gate and its application in pulse train-controlled Buck converter. *Int J Circ Theory Appl*. 2022;50(6):2190-2209. doi:10.1002/cta.3260
11. An H, Ehsan MA, Zhou Z, Shen F, Yi Y. Monolithic 3D neuromorphic computing system with hybrid CMOS and memristor based synapses and neurons. *Integration Journal*. 2019;65:273-281. doi:10.1016/j.vlsi.2017.10.009
12. Wang X, Chen Y. Spintronic memristor devices and application. in Proc. Des. Autom. Text Eur. Conf. Exhibit, 2010, 667–672. 10.1109/DATE.2010.5457118.
13. Cifarelli A, Parisini A, Iannotta S, Berzina T. Organic memristor device based on pectin as a solid polyelectrolyte. *Microelectron Eng*. 2018;185:55-60. doi:10.1016/j.mee.2017.11.001
14. Lu W, Li W, Wen D. Soybean-based memristor for multilevel data storage and emulation of synaptic behaviour. *Microelectron Eng*. 2023;267:111911. doi:10.1016/j.mee.2022.111911
15. Jo SH, Chang T, Ebong I, Bhadivya BB, Mazumder P, Lu W. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett*. 2010;10(4):1297-1301. doi:10.1021/nl904092h
16. Pershin YV, Ventra MD. Practical approach to programable analog circuits with memristors. *IEEE Trans Circuits Syst I Reg Paper*. 2010;57(8):1857-1864. doi:10.1109/TCSI.2009.2038539
17. Zdeněk K, Biolek D, Biolková V. Hybrid modelling and emulation of mem-systems. *Int J Numer Model Electron Netw Dev Fields*. 2012;25(3):216-225. doi:10.1002/jnm.825
18. Fitch AL, Iu HH, Wang XY, Sreeram V, Qi WG. Realization of an analog model of memristor based on light dependent resistor. *2012 IEEE International Symposium on Circuits and Systems (ISCAS), Seoul, Korea* (South). 2012: 1139–1142. 10.1109/ISCAS.2012.6271433.
19. Abuelama'atti MT, Khalifa ZJ. A new floating memristor emulator and its application in frequency-to-voltage conversion. *Anal Integr Circuit Signal Process*. 2016;86(1):141-147. doi:10.1007/s10470-015-0660-6
20. Hussein AI, Fouda ME. A simple MOS realization of current controlled memristor emulator, *2013 25th International Conference on Microelectronics (ICM)*, Beirut, Lebanon, 2013: 1–4. 10.1109/ICM.2013.6734969.
21. Sozen H, Cam U. Electronically tunable memristor emulator circuit. *Analog Integr Circuits Signal Process*. 2016;89(3):655-663. doi:10.1007/s10470-016-0785-2
22. Valsa J, Biolek D, Biolek Z. An analogue model of the memristor. *Int J Numer Model Electron Netw Dev Fields*. 2011;24(4):400-408. doi:10.1002/jnm.786
23. Lopez CS, Aguilar MAC, Montero CM. A 16 Hz-160 kHz memristor emulator circuit. *Int J Electron Commun*. 2015;69(9):1208-1219. doi:10.1016/j.aeue.2015.05.003
24. Lopez CS, Aguilar MAC, Montero CM. A floating analog memristor emulator circuit. *IEEE Trans Circuits Syst II*. 2014;61(5):309-313. doi:10.1109/TCSII.2014.2312806
25. Yesil A, Babacan Y, Kacar F. A new DDCC based memristor emulator circuit and its applications. *Microelectron J*. 2014;45(3):282-287. doi:10.1016/j.mejo.2014.01.011
26. Ayten UE, Minaei S, Sagbas M. Memristor emulator circuits using single CBTA. *Int J Electron Commun*. 2017;82:109-118. doi:10.1016/j.aeue.2017.08.008
27. Sánchez-López C, Aguila-Cuapio LE. 860 kHz grounded memristor emulator circuit. *AEU Int J Electron Commun*. 2017;73:23-33. doi:10.1016/j.aeue.2016.12.015
28. Babacan Y, Yesil A, Kacar F. Memristor emulator with tunable characteristic and its experimental results. *AEU Int J Electron Commun*. 2017;81:99-104. doi:10.1016/j.aeue.2017.07.012
29. Petrovic PB. Floating incremental/decremental flux-controlled memristor emulator circuit based on single VDTA. *Analog Integr Circuits Signal Process*. 2018;96(3):417-433. doi:10.1007/s10470-018-1177-6

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

30. Xie X, Zou L, Wen S, Zeng Z, Huang T. A flux-controlled logarithmic memristor model and emulator. *Circuits Syst Signal Process*. 2019; 38(4):1452-1465. doi:10.1007/s00034-018-0926-1
31. Wei L, Wang FQ, Ma XK. A unified cubic flux-controlled memristor: theoretical analysis, simulation and circuit experiment. *Int J Numer Model Electron Netw Dev Fields*. 2015;28(3):335-345. doi:10.1002/jnm.2009
32. Pal I, Kumar V, Aishwarya N, Nayak A, Islam A. A VDTA-based robust electronically tunable memristor emulator circuit. *Analog Integr Circuits Signal Process*. 2020;104(1):47-59. doi:10.1007/s10470-019-01575-y
33. Kanyal G, Kumar P, Paul SK, Kumar A. OTA based high frequency tunable resistorless grounded and floating memristor emulators. *Int J Electron Commun*. 2018;92:124-145. doi:10.1016/j.aeue.2018.05.027
34. Arundeepakvel R, Khatter P, Pandey N, Minaei S. A novel design for voltage inverting metamutator and its applications. *Microelectron J*. 2021;113:105096. doi:10.1016/j.mejo.2021.105096
35. Bhardwaj K, Srivastava M. New multiplier-less compact tunable charge controlled memelement emulator using grounded passive elements. *Circuits Syst Signal Process*. 2021;30(5):1-37. doi:10.1007/s00034-021-01895-3
36. Fatih G, Babacan Y. A novel OTA-based circuit model corroborated by an experimental semiconductor memristor. *Microelectron Eng*. 2018;194:56-60. doi:10.1016/j.mee.2018.03.012
37. Vista J, Ranjan A. Flux controlled floating memristor employing VDTA: incremental or decremental operation. *IEEE Trans Comput Aided des Integr Circuits Syst*. 2021;40(2):364-372. doi:10.1109/TCAD.2020.2999919
38. Kumar N, Kumar M, Pandey N. Single DXCCTA based charge controlled floating incremental/decremental memristor emulator, 2022 *8th International Conference on Signal Processing and Communication* (ICSC), Noida, India, (2022) 663–668. 10.1109/ICSC56524.2022.10009249.
39. Ranjan RK, Rani N, Pal R, Paul SK, Kanyal G. Single CCTA based high frequency floating and grounded type of incremental/decremental memristor emulator and its application. *Microelectron J*. 2017;60:119-128. doi:10.1016/j.mejo.2016.12.004
40. Yadav N, Rai SK, Pandey R. New grounded and floating memristor emulators using OTA and CDBA. *Int J Circ Theory Appl*. 2020;48(7): 1154-1179. doi:10.1002/cta.2774
41. Ranjan RK, Raj N, Bhuwal N, Khateb F. Single DVCCTA based high frequency incremental/decremental memristor emulator and its application. *Int J Electron Commun*. 2017;82:177-190. doi:10.1016/j.aeue.2017.07.039
42. Kumar P, Kausik BK, Ranjan RK. A novel second-generation current conveyor (CCII)-based high frequency memristor model. *Microelectron Eng J*. 2023;271-272:111938. doi:10.1016/j.mee.2023.111938
43. Raj N, Ranjan RK, Khateb F. Flux-controlled memristor emulator and its experimental result. *IEEE Trans Very Large-Scale Integr (VLSI) Syst*. 2020;28(4):1050-1061. doi:10.1109/TVLSI.2020.2966292
44. Sedra AS, Smith KC, Chandorkar AN. *Microelectronics Circuits Theory and Applications*. Seventh ed. Oxford University Press; 2014.
45. Adhikari SP, Sah MP, Kim H, Chua LO. Three fingerprints of memristor. *IEEE Trans Circuits Syst I Reg Pap*. 2013;60(11):3008-3021. doi: 10.1109/TCSI.2013.2256171

How to cite this article: Kumar N, Kumar M, Pandey N, Minaei S. Second generation current conveyor based capacitorless floating memristor emulator. *Int J Circ Theor Appl*. 2024;1-20. doi:10.1002/cta.4175

1097007x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/cta.4175 by Delhi Technological University, Wiley Online Library on [05/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2024 5th International Conference for Emerging Technology (INCET)
Karnataka, India. May 24-26, 2024

Sentiment Analysis of Unified Payments Interface (UPI) Payment App Review using Machine Learning Algorithm

Kunal Jaluthria
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
kunaljaluthria@gmail.com

Karandeep Singh
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
kdeepsingh2002@gmail.com

Manik Sharma
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
themannu07@gmail.com

Rajni Jindal
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
rajnijindal@dce.ac.in

2024 5th International Conference for Emerging Technology (INCET) | 979-8-3503-6115-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/INCET61516.2024.10593356

***Abstract*—The paper "Sentiment Analysis of UPI Application Reviews using Machine Learning Algorithm" uses natural language processing (NLP) and machine learning to analyse user evaluations of Unified Payments Interface (UPI) applications for sentiments.The objective is to conduct a comparative analysis of different algorithms to better understand the user reviews of UPI applications and provide useful data for these digital platforms' strategic enhancements. Preprocessing is done on datasets, which includes text transformation with techniques like Word2Vec and Bag of Words. These techniques transform raw text into numerical vectors that are used by machine learning algorithms. Reviews are categorised into positive and negative sentiments using sentiment analysis algorithms. For sentiment classification, algorithms like C 4.5 (J48), Naive Bayes, and LSTM are used. Their effectiveness is assessed through comparative analysis. 90.02% was the highest accuracy recorded for sentiment classification of user reviews on Paytm. LSTM in conjunction with the word2vec skip-gram training model resulted in the attainment of this prediction. Results provide perceptions into user sentiments, facilitating purposeful enhancements in UPI applications that boost user contentment and adhere to changing digital payment requirements.**

***Keywords*—Sentiment Analysis, Natural Language Processing, Machine learning, Neural Networks, UPI, Word Embedding, Word2Vec, Naive Bayes, C4.5, LSTM**

I. Introduction

According to recent statistics, there has been a notable increase in UPI transactions in the recent past. These figures show a significant rise, with the UPI transaction volume showing an incredible 62% increase. In particular, the volume increased from 31.95 billion transactions in 2022's first half to a remarkable 51.91 billion transactions in 2023's same time[10]. User sentiments are crucial in determining the competitiveness and profitability of platforms in the dynamic world of mobile payment apps. To improve user experiences and effectively resolve problems, developers and stakeholders must have a sophisticated understanding of consumers' sentiments regarding particular applications. This study's detail makes it an invaluable addition to the ongoing efforts to enhance user experiences in the rapidly expanding field of digital payments.

Sentiment analysis, sometimes referred to as opinion mining, is the process by which machines accurately detect the emotions expressed in text, such as reviews or posts. Algorithms for sentiment analysis are essential for deriving insightful information from textual input. The analysis of feedback from users presents challenges since machines have difficulty identifying sarcasm. By using word2vec, this study aims to improve accuracy over existing sentiment analysis algorithms that have been used in the past.

This study uses various algorithms, including C 4.5, Naive Bayes, and the deep learning algorithm LSTM. To better align these digital payment platforms with user expectations and promote overall user satisfaction in the competitive world of mobile payment applications, the study's main goals are to conduct a comparative analysis to determine which one yields the highest accuracy in this particular situation and to offer insightful analysis that could direct future improvements. This work aims to investigate the potential applications of word embeddings produced by the word2vec algorithm for sentiment classification. The main idea is to use word embeddings to leverage sentiment-classifying features instead of manually constructing them.

This study's main contributions are as follows:

- A labelled dataset of reviews of Paytm applications that have been classified as either positive or negative.
- Building models with different algorithms designed

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:48:39 UTC from IEEE Xplore. Restrictions apply.

specifically for user review sentiment analysis.
- Determining the ideal Word2Vec parameters, including vector dimension, model architecture, and evaluation technique.

This paper's remaining sections are organised as follows:

Section 2 examines the existing literature of research on sentiment analysis, the Word2Vec method, and other noteworthy sentiment analysis models. The technique and the preprocessing- ing methods used on the dataset are described in Section 3-7. Section 8 presents the findings of the experiment. The paper is concluded in Section 9, where our results are summarised and compared to those of other experiments. Section 10 concludes by outlining possible directions for further research.

II. Literature Review

Baid, Gupta, and Chaplot's research [1] focuses on sentiment analysis applied to movie reviews using machine learning techniques. The research investigates the efficacy of different machine learning models in extracting sentiment from textual movie reviews.The research compares the KNN, Naive Bayes, and Random Forest algorithms for sentiment analysis of reviews. The paper's focus on sentiment analysis within the context of movie reviews is noteworthy, as it provides insights into the use of machine learning algorithms for sentiment classification. The study's significance is enhanced by the actual application of sentiment analysis in the context of movie reviews. However, as it mentions certain other algorithms also need to be tested.

Ngoc et al. in their paper [2] focus on emotional classification in English text, employing the C 4.5 algorithm. The paper investigates the effectiveness of C 4.5 in detecting emotional nuances in textual content. The dataset utilised in the study is fairly well documented and contains thousands of examples. However, the study lacks some critical preprocessing procedures that could improve the study's accuracy and efficiency. Each technique contributes to the detailed refinement of our dataset.

Vanaja and Belwal in their paper [3] present a method for aspect-level sentiment analysis, aiming to capture the intricacies of sentiment related to specific aspects of products or services. The study uses two machine learning approaches, Naive Bayes and Support Vector Machine, to address an important component of sentiment analysis in the realm of e-commerce data. It discusses the significance of evaluating e-commerce data for online retailers. The future work of the study focuses on using the C 4.5 algorithm and comparing it with others.

Saputra, Yulia, and Sholikha in their paper [4] investigate sentiment in Indonesian tweets related to e-money during the pandemic, employing both C 4.5 and Naive Bayes algorithms for comparison. They present statistics on the rise in e-money payment volume. They carried out a classification of negative and positive opinions related to e-money in Indonesia. The paper's contextual relevance is commendable, as it addresses sentiment analysis within the unusual conditions of a pandemic. However, sentiment analysis requires a high number of instances due to the sheer amount of words and phrases, and the dataset used in this classification study is not too vast.

In their research [5], Sholihat et al. use the Naive Bayes algorithm to analyse the sentiment of Twitter posts about illegal investments with the hashtags illegal investment and crazy rich. It was intended to address public concerns about illegal investments. According to the paper, their study performed more accurately than the prior studies.

Astuti, Wahyuni, Maulina, and Sidiq in their paper [6] explore sentiment analysis in the context of data leakage, employing the Naive Bayes based on machine learning approach.The research predicts sentiment patterns on public data leaks. The research addresses a crucial aspect of sentiment analysis in the domain of data security and leakage. The research clearly explains the methodology along with the algorithm used. However only Naive Bayes Algorithm was tested, it is required to test a few other algorithms as well.

Acosta et al. in their paper [7] explore the application of word2vec in sentiment analysis of Twitter messages. Drawing from existing literature, the study addresses challenges unique to analysing sentiments in tweets and evaluates the effectiveness of word2vec in capturing contextual information. By integrating word embeddings with Naive Bayes, Support Vector Machines and Logistical Regression machine learning algorithms, the paper contributes to advancing sentiment analysis techniques for social media data mining.

Rizka Putri Nawangsari et al. in their paper [8] explore the application of Word2Vec in the domain of sentiment analysis, particularly focusing on hotel reviews in the Indonesian language. They address the challenges associated with hand-crafted feature engineering, emphasising its labour-intensive nature and high cost. Instead, they propose Word2Vec as an alternative, leveraging its ability to generate vector representations of words, thereby facilitating the automation of feature extraction for sentiment analysis tasks. The study provides insights into achieving optimal accuracy in sentiment analysis of hotel reviews.

Putra Fissabil Muhammad et al. in their paper[9] explore sentiment analysis in Indonesian hotel reviews using Word2Vec and Long Short-Term Memory (LSTM) models. Online Travel Agents (OTAs) often include a review feature, enabling customers to provide feedback on their experiences with various services, including hotels. Sentiment analysis of these reviews aids both users in making informed decisions and service providers in evaluating customer satisfaction. The study employs LSTM, a type of Recurrent Neural Network (RNN) designed to retain information over long sequences. Results indicate that the Skip-gram architecture of Word2Vec outperforms CBOW.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:48:39 UTC from IEEE Xplore. Restrictions apply.

III. Methodology

This paper presents a methodological framework for the systematic analysis of user reviews utilising machine learning algorithms. The research methodology employs a quantitative approach, progressing through distinct phases to extract valuable insights from user-generated content. To study the sentiment of user reviews, the experiments were conducted on a publicly available dataset (www.kaggle.com) of Google Play Store app reviews relating to users' experiences with the UPI payment application(Paytm). Subsequently, datasets undergo meticulous labelling procedures to enhance accuracy and relevance. Text preprocessing techniques are then applied to refine the textual data, optimising its utility for subsequent analysis. To make the datasets compatible with machine learning algorithms, they are preprocessed. This involves turning raw text into numerical vectors using methods like Bag of Words and Word2Vec. The main components of the methodology are the algorithms Naive Bayes, C 4.5, and Long Short-Term Memory (LSTM), which are intended to identify complex patterns and relationships in the dataset. Ultimately, extensive testing is carried out to assess the model's effectiveness and ensure the reliability of the findings. This methodological framework offers a systematic and thorough approach to user review analysis, combining advanced technological tools with analytical precision to derive robust conclusions.



Fig. 1. Methodology

IV. Data Collection

For this study, we used the open online platform Kaggle to gather useful data, with a focus on user reviews of well-known UPI payment services like Paytm. This large dataset serves as the foundation for our sentiment analysis study and allows us to gain insights into the opinions and experiences of users of these digital payment services. By using this strategy, we guarantee a thorough investigation of customer opinions regarding Paytm. Based on the review score, the gathered data is carefully labelled, with entries being classified as positive ($score > 3$) or negative ($score <= 3$)). The dataset has a set of attributes before it is labelled, which are described below and help with further analysis. The dataset used in this study had 68952 data points in total, 56520 of which had positive labels and 12432 of which had negative labels.

V. Data Preprocessing

Within the field of sentiment analysis, data preprocessing plays a crucial role as the foundational phase towards transforming unprocessed data into a format that is consistent and allows for insightful analysis. We carefully carried out several preprocessing procedures as part of our study, which

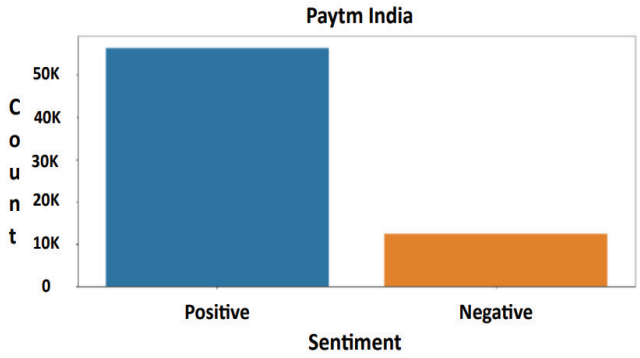


Fig. 2. Collected Paytm Review Data

TABLE I
Example of Collected Paytm Data

Attribute	Values
Num	2
reviewID	gp:AOqpTOGse_Ua7yhZRYP0bvEWZ4 DqJlOA JWLQC1 376sQ...
userName	Rajendra Kumar
content	My paytm wallet app is good
score	4
At	2021-11-18 18:41:35
replyContent	Hi, Rajendra ! We'd like to take a closer look a...

was centred on identifying sentiments in customer reviews of Paytm. Each step helped to refine our dataset subtly. This is a thorough examination of every preprocessing step:

A. Data cleaning and attribute selection

The first step involved clearing the data and eliminating any unnecessary attributes, including the review ID, username, user image, reply count, and review creation version. By concentrating the information down to its essential components—content/review and score, for example—we focused our research on the critical elements that significantly contribute to sentiment classification.

B. Handling missing values

We took the vital action of removing items without accompanying reviews to preserve the integrity of our dataset. This simplifies the data and improves the precision of the analyses that follow by eliminating instances that lack meaningful textual information.

C. Converting to Lowercase

To ensure consistency in analyses that follow, the text must be standardised by changing all of the characters to lowercase. To avoid any potential differences, this step ensures that words are handled consistently, regardless of their original case.

D. Removing Punctuations, Numbers, and Special Characters

Eliminating unnecessary components like punctuation, digits, and special characters is the next level of refining. By

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:48:39 UTC from IEEE Xplore. Restrictions apply.

concentrating on the key terms that support the meaning of sentiment, this method streamlines the text.

E. Removing URLs

Removing URLs from the text is essential to keep focus on the language used in reviews. If they are there, URLs could add noise to the dataset and cannot assist with sentiment expression.

F. Replacing Slang Words

Full forms of slang words were used in place of their informal, frequently shortened counterparts. By standardising the language, this transformation helps ensure that the sentiment analysis accurately captures the full range of user expressions.

G. Removing Stop Words

Stop words, which are frequently used terms with no semantic meaning, were omitted. This stage refines the dataset by removing terms such as ”the,” ”a,” ”an,” and ”in” that make no substantial contribution to sentiment expression.

H. Tokenization

Breaking down the text into distinct tokens, or words, is critical for future analysis. Tokenization lays the framework for interpreting sentences' semantic structure, allowing for a more detailed study of sentiments.

I. Stemming

The last phase is stemming, which involves reducing words to their base or root form. This ensures that different word variations—such as ”easy” and ”ease,” or ”running” and ”run”—are treated as the same thing, improving sentiment analysis algorithms' effectiveness.

VI. Text transformation

A key step in the advancement of our sentiment analysis study was applying the ”StringToWordVector” transformation to our textual data. This NLP approach converted strings of text into useful numerical representations known as word vectors. This transformation helps to prepare the data for later machine learning analysis. It simplifies the conversion of textual data into organised word vectors, setting the groundwork for more advanced analytics. Following the transformation, each document was represented as a numerical vector. Each member of the vector represents a specific word from the lexicon, and its value reflects the frequency or presence of that term in the document.

Furthermore, Word2Vec is a popular approach for creating word vectors. Word vectors are numerical representations of words that convey semantic meaning dependent on the context in which they appear. Word2Vec learns vector representations of words by training neural networks on massive text corpora. Using techniques such as Word2Vec, we may convert textual data into organised word vectors, setting the groundwork for advanced data analysis. Word2Vec helps capture the semantic meaning of words, allowing sentiment analysis models to better comprehend the context and sentiment expressed in the text.

In addition to its popularity, Word2Vec's Skip-gram design is noteworthy for its effectiveness in creating word vectors. Word2Vec uses two models: the Skip-gram architecture and the Continuous Bag of Words (CBOW). Unlike CBOW, which predicts the current word based on its context, Skip-gram predicts context words using the current word. This design is especially beneficial for capturing the contextual and semantic meaning of words in more detail. By taking the surrounding words as context, Skip-gram can effectively learn the links between words and their contexts, leading to richer and more advanced word embeddings. Skip-gram's ability to capture subtle semantic links between words helps improve understanding of sentiment represented in textual data, thereby increasing sentiment analysis model performance. Thus, introducing Skip-gram architecture into the Word2Vec training process enhances word embeddings with useful contextual information, allowing for more accurate sentiment analysis.

The vectors were then used to train various classifiers to determine which classifier produced the most accurate classifications. The findings were documented and explained further in the results section.

In this study, the Pandas library was used to load and manipulate the dataset, which included activities like removing extraneous columns, handling missing values, and preparing the data for analysis. NumPy also plays an important role in facilitating numerical computations and array operations during the data preprocessing stage. Matplotlib and Seaborn were used for data visualisation tasks, allowing for the production of useful plots and visualisations to investigate the distribution of sentiment categories in the dataset. Regular expressions from the re-library improved text processing by removing punctuation, special characters, and URLs. NLTK provided significant tools for natural language processing, including tokenization, stopword elimination, and lemmatization. Scikit-learn was used to divide the dataset into training and testing sets, while Keras was the major framework for developing and training deep learning models for sentiment analysis. Finally, Gensim was used to train word embedding models like Word2Vec, which helped convert textual data into numerical vectors while keeping semantic links between words.

VII. Classification

We examined the performance of three different classifiers: Naive Bayes, C4.5 decision trees, and Long Short-Term Memory (LSTM) neural networks. Each classifier was evaluated on its ability to identify sentiment in text data. Furthermore, the Skip-gram word embedding technique was integrated with the Word2Vec algorithm to improve the semantic representation of words. The dataset was divided into 80% training and 20% testing to properly evaluate the classifiers' performance. The next part describes the experimental results and evaluation criteria used to assess the classifiers' effectiveness in sentiment analysis tasks.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:48:39 UTC from IEEE Xplore. Restrictions apply.

VIII. Results

Our study evaluates the accuracy of the Naive Bayes and C4.5 algorithms when applied to text data modified using a bag-of-words technique. We also trained the LSTM and prior classifiers using Skip-gram, which is integrated with the Word2Vec algorithm, as shown in Table 2.

TABLE II
ACCURACY PERCENTAGE BY CLASSIFIER AND TEXT TRANSFORMATION

Classifier	Text transformation	Accuracy
Naive Bayes	Bag of words	84.43%
Naive Bayes	Word2Vec (Skipgram architecture)	80.97%
C4.5	Bag of words	89.47%
C4.5	Word2Vec (Skipgram architecture)	87.46%
LSTM	Word2Vec (Pre-trained embedding)	88.04%
LSTM	Word2Vec (Custom trained embedding)	90.02%

Our results demonstrate that the Naive Bayes and C4.5 with bag of words text transformations have an accuracy of 84.43% and 89.47%, respectively. Also, according to the analysis, the C 4.5 approach takes much more time than the naive Bayes algorithm. Naive Bayes and C4.5, when employed with Word2Vec (Skip-gram), achieve an accuracy of 80.26% and 87.46%, respectively. Furthermore, the highest accuracy was found in LSTM using Word2Vec at 90.02% when the dropout value, learning rate, and vector size were 0.2, 0.001, and 100, respectively. When used with the LSTM classifier, Google's pre-trained Word2Vec embedding achieved an accuracy of 88.04%. In this use scenario, our custom-trained embedding outperformed the pre-trained word2vec embedding.

TABLE III
CLASSIFICATION REPORT (AVERAGE VALUES)

Algorithm	Precision	Recall	F-measure
Naive Bayes & Bag of words	0.84	0.85	0.84
Naive Bayes & Word2Vec	0.84	0.81	0.82
C4.5 & Bag of words	0.89	0.9	0.88
C4.5 & Word2Vec	0.87	0.88	0.87
LSTM & Word2Vec (Pre-trained)	0.89	0.97	0.93
LSTM & Word2Vec (Custom)	0.91	0.97	0.94

Determining the sentiment of reviews is important in many different contexts. It enables the creation of intelligent systems capable of providing customers with detailed reviews of UPI payment apps, products, and services. This removes the need for users to wade through individual evaluations, allowing them to make decisions based on the insights supplied by these intelligent systems.

IX. Conclusion

In conclusion, our study thoroughly analysed the performance of various machine learning algorithms in sentiment analysis of user reviews for the Unified Payments Interface (UPI) app (Paytm). This study tried to establish whether the Word2Vec technique, which generates word embeddings, might be used to classify sentiment. Using word embeddings would save the analysts from having to manually build stylometry-based features to appropriately classify data.

Our study found that using C4.5 and Na¨ıve Bayes algorithms with a bag-of-words text transformation approach resulted in higher accuracies (89.47% and 84.43%, respectively) compared to previous studies. Using Word2Vec embeddings, especially Skip-gram, with C4.5 and Na¨ıve Bayes achieved accuracy of 87.46% and 80.26%, respectively. Although the accuracy remained nearly equal, Bag of Words coupled with Naive Bayes and c4.5 outperformed Word2Vec paired with these algorithms. This could be attributed to the dataset's restricted contextual diversity. Notably, the LSTM model outperformed all other techniques, obtaining a maximum accuracy of 90.02% when combined with Word2Vec Skip-gram embeddings. Furthermore, we found that our custom-trained embedding outperformed pre-trained Word2Vec embeddings from Google.

Our findings highlight the importance of using advanced approaches like Word2Vec embeddings and LSTM models for sentiment analysis tasks since they provide important insights into user sentiments surrounding UPI applications. These findings can help digital platforms make strategic modifications to boost customer satisfaction and adapt to changing digital payment needs. Overall, our comparative analysis helps to gain a better knowledge of sentiment analysis approaches and their implications for improving user experiences in digital payment ecosystems.

X. Future scope

In future research on sentiment analysis for Unified Payments Interface (UPI) applications, incorporating transformer-based models such as BERT could greatly improve performance. Exploring alternatives such as FastText[11] embedding and hybrid techniques provides exciting opportunities for deeper insights and greater accuracy. FastText's subword information may improve sentiment representation, whilst hybrid techniques may address individual limitations. Integrating these improvements provides a complete strategy for refining sentiment analysis, which is critical for making informed decisions and improving user experiences in digital payment networks. Such initiatives pave the way for more effective sentiment analysis that is suited to the organisation's specific needs.

References

[1] P. Baid, A. Gupta, and N. Chaplot, "Sentiment analysis of movie reviews using machine learning techniques," Int. J. Comput. Appl., vol. 179, no. 7, pp. 45–49, 2017.

[2] P. V. Ngoc, C. V. T. Ngoc, T. V. T. Ngoc, and D. N. Duy, "A C 4.5 algorithm for english emotional classification," Evol. Syst., vol. 10, no. 3, pp. 425–451, 2019.

[3] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," in 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 2018.

[4] W. S. J. Saputra, P. Eva Yulia, and Z. E. Sholikha, "C 4.5 and naive Bayes for sentiment analysis Indonesian tweet on E-money user during pandemic," in 2020 6th Information Technology International Seminar (ITIS), 2020.

[5] A. Sholihat, F. Bei, R. Ainaya, F. Sembiring, and A. Lattu, "Twitter tweet: Sentiment analysis on illegal investment using naive Bayes algorithm," in 2022 IEEE 8th International Conference on Computing, Engineering and Design (ICCED), 2022.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:48:39 UTC from IEEE Xplore. Restrictions apply.

[6] Y. Astuti, S. N. Wahyuni, D. Maulina, and F. M. Sidiq, "The data leakage sentiment analysis using naive Bayes algorithm based on machine learning approach," in 2022 5th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2022.
[7] J. Acosta, N. Lamaute, M. Luo, E. Finkelstein, and A. Cotoranu, "Sentiment Analysis of Twitter MessagesUsing Word2Vec" in Student-Faculty Research Day, CSIS, Pace University, 2017
[8] R. P. Nawangsari, R. Kusumaningrum, and A. Wibowo, "Word2Vec for Indonesian sentiment analysis towards hotel reviews: An evaluation study," Procedia Comput. Sci., vol. 157, pp. 360–366, 2019.
[9] P. F. Muhammad, R. Kusumaningrum, and A. Wibowo, "Sentiment analysis using Word2vec and long short-term memory (LSTM) for Indonesian hotel reviews," Procedia Comput. Sci., vol. 179, pp. 728–735, 2021.
[10] Outlook Money and Outlook Money, "India's UPI Transaction Volume Surged 62% YoY In HI 2023—Know Trends in Digital Payments," Outlook Business and Money, Sep. 27, 2023. [Online]. Available: https://business.outlookindia.com/banking/indias-upi-transaction-volume-surged-62-yoy-in-hi-2023know-trends-in-digital-payments
[11] S. Khomsah, R. D. Ramadhani, and S. Wijaya, "The accuracy comparison between Word2Vec and FastText on sentiment analysis of hotel reviews," J. RESTI (Rekayasa Sist. Dan Teknol. Inf.), vol. 6, no. 3, pp. 352–358, 2022.
[12] Everydaycodings, "EDA On UPI Payment Apps review," Kaggle, Nov. 30, 2021. [Online]. Available: https://www.kaggle.com/code/everydaycodings/eda-on-upi-payment-apps-review

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on August 06,2024 at 05:48:39 UTC from IEEE Xplore. Restrictions apply.

Shapley Explainable Deep Learning based Knowledge Distillation Framework for Student's Performance Prediction

1st Rahul
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
rahul@dtu.ac.in

2nd Rahul Katarya
Department of Computer Science and Engineering
Delhi Technological University
Delhi, India
rahuldtu@gmail.com

2024 Second International Conference on Data Science and Information System (ICDSIS) | 979-8-3503-8166-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDSIS61070.2024.10594386

***Abstract*—Educational institutions are constantly striving to improve the services they provide; their goal is to have the best possible teaching staff, increase the quality of teaching, and improve their students' academic achievement. One of the initial steps taken by educational institutions was to establish Virtual Learning Environments (VLEs). This work uses a deep Explainable distillation model on a set of unique handcrafted variables collected from VLE clickstream data to forecast at-risk students and provide strategies for early intervention in such circumstances. Initially, the input data are gathered from publicly available source and pre-processing is done to enhance the quality of inputs using Missing Value Imputation. Then, in order to handle the unbalanced data, augmentation is performed by a new Improved Synthetic Minority Oversampling Technique (ISMOTE). After augmenting the data, the student's academic performance is predicted by proposing a novel Explainable Deep Learning based Knowledge Distillation Framework (EDL_KDF). Furthermore, the proposed model can be explained using Shapley Additive Explanations (SHAP)-based reasoning. Educational institutions can evaluate, audit, and validate explanations for such relevant features more easily than explanations from DL models. The proposed framework attained 99.6% % accuracy and 99.21% of precision over the Open University Learning Analytics Dataset (OULAD). Thus, the proposed study effectively predicts the students' performance compared to other existing models.**

Keywords—deep learning, student performance prediction, knowledge distillation, shapley additive explanation, improved SMOTE, OULAD dataset, resnet-101.

I. INTRODUCTION

Sudden advancements in enhanced technology-based learning platforms have provide a tremendous rise in online education related data for attaining plenteous educational repositories [1]. In recent years, it is noticed that e-learning systems have afford great flexibility and interactivity because which can easily use by the learners at any time in varied locations and devices [2]. Mainly, such systems are employed for blended and online universities by instructors to distribute, organize educational information to learners or students [3]. Nevertheless, the instructors in online platforms does not analyze academic performance and behavior of students. On the other hand, face-to-face instructors has the ability to attain the knowledge of student's behavior and easy to predict their academic performance [4]. Online instructors also can communicate with students in virtual manner but it face difficult to analyze the performance of each students [5].

Nowadays, educational institutions are becoming a large competitive environment because of enhancing people's interest in several courses [6]. Also, the need of assuring that the utilized resources is highly important because it can enhance the experience of student learning and increase esteem factors like student's performance and retention. Analyzing student performance is necessary as it aids in designing a student support strategy and enhance learning and teaching practice [7]. However, conducting a deep analysis of student academic efficacy is a challenging problem and some factors like social, economic, cultural, demographic and academic background may impacts the student's academic results. Developing an effective student academic performance factor demands a deep data analysis which can attain via Machine Learning (ML) and Deep Learning (DL) models [8]. Such models has the capability to attain more informative features from the input samples and can predicts the exact outcomes. Some of the preceding methods that used for analyzing student's academic performance are artificial neural network (ANN), deep neural networks (DNN), support vector machine (SVM) etc.

A. Motivation and Problem statement

Learning analytics based on Artificial Intelligence (AI) is one of the promising technology which highly support educators because which can collects, analyze and reports data regarding learners. However, predicting student's performance in online platforms is becoming a challenging issue because of the complex inputs, overfitting problems, reduced learning ability, etc. Recently, several studies have developed different models for making student's performance prediction. But, the performance is limited due to its inefficiency in prediction task. Nowadays, DL models have gained more attention in prediction and classification tasks due to its efficient characteristics. But, the complex structures of DL mechanisms impacts the network performance as well as it produced high computational complexity. Thus, to avoid this kind of issues, the proposed study plans to utilize an effective knowledge distillation-based DL models for enabling prediction process. The major objectives of the proposed work are,

- To enhance the quality of inputs by filling the missing values in the initial pre-processing step and thereby smoothening the further stages.
- To reduce the data imbalance issue by augmenting the data samples using a new Improved Synthetic Minority Oversampling Technique (ISMOTE).
- To propose a novel Explainable Deep Learning based Knowledge Distillation Framework (EDL_KDF) for predicting student's academic performance with less computational complexity and high accuracy.

The paper is structured as five different sections. Section 2 contained the review of the related works. Section 3

979-8-3503-8166-5/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

contained the detail explanation about the proposed methodology. Section 4 contained the result analysis and discussion and section consists of overall conclusion of the conference paper with future direction and references.

II. Related Works

Waheed et al. [9] have designed deep ANN for learning the predictive function of student's performance using features from the VLE. The simulation results shows that the developed model attains better classification accuracy as 84% to 93%. Giannakas et al. [10] have developed a DNN model for enabling predictions on academic team performance. Here, the utilized DNN model is trained by using varied activation functions and optimizers. For interpreting the developed framework, SHAP method is used. Also, the necessary features that involves positive and negative impacts on its resultant prediction process are extracted. DNN attained 86.57% and 80.76% for Adagrad and Adadelta optimizers correspondingly.

Asselman et al. [11] have used ML based ensembled algorithms for predicting students' performance. The study developed a performance factors analysis (PFA) method based on varied models like adaboost, XGBoost and random forest. The model attained 78.75% of accuracy over three different datasets. Rodríguez-Hernández et al. [12] have presented a systematic process for implementing ANN to analyze academic performance of students in their higher studies. This existing study used 162, 030 number of student data including males and females from the public and private universities in Colombia. The ANN model attained 82% of high accuracy. Kusumawardani et al. [13] suggested the DL based transformer encoder for student performance prediction using OULAD dataset. The encoder attained 83.17% of accuracy. From the analysis the existing model have several drawbacks such as less accuracy, high computational complexity, overfitting and dataset class imbalance issues occurred during the prediction.

III. Proposed Methodology

Predicting the academic performance of the student in the e-learning platform is more essential. In recent development of DL technology several methods are introduced but it have several limitation such as more computational complexity, overfitting problem etc. So that the proposed framework is developed to attain accurate student performance prediction with less computational complexity. Figure 1 shows the architecture of the proposed prediction model.

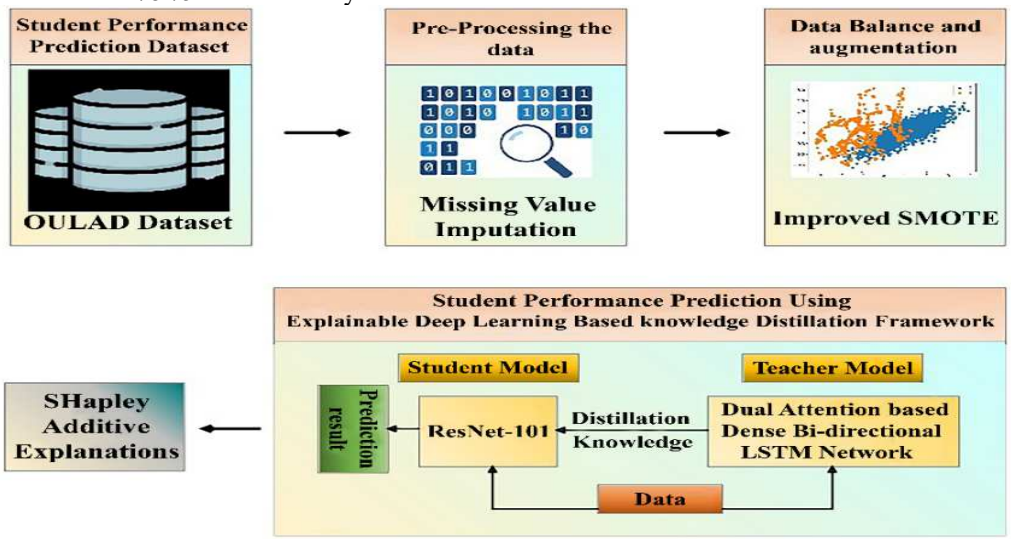


Fig. 1. Architecture of proposed framework

Initially, the input data is pre-processed using the Missing Value Imputation then data balance and data augmentation done using the ISMOTE that improve the quality of the dataset. After that, EDL_KDF is introduced for student performance prediction, which contained the DAbDBi-LSTM teacher model and ResNet-101 as the student model. Finally the predicted result is analyzed using SHAP to know the impact of each features presented in the dataset.

A. Pre-processing the input data using Missing value imputation

In general, due to some issues few values are missing in the dataset that can be replaced through the mean value of the neighboring non-missing values. The mean value is estimated by the equation (1).

$$V_i = \frac{V_{i-1}+V_{i+1}}{a} i \in C \quad (1)$$

Where, missing value is signified as V_i, previous value of the missing value is signified as V_{i-1} like that, the next value followed by the missing value is denoted as V_{i+1}. The natural number is represented as $C = 1,2,3,\ldots$. Through these techniques, the missed values in the dataset is effectively replaced, so that it enhanced the quality of the input data.

B. Data balance using Improved Synthetic Minority Oversampling Technique

ISMOTE intends to generate clear synthetic samples for the minority class of imbalanced dataset to data augmentation through SMOTE generalization in addition to accurately establishing the boundaries. Process of ISMOTE is explained as follow. Initially, synthetic instance is created through employing the SMOTE techniques that expressed in equation (2).

$$S = 2 * (u - v) + v \quad (2)$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

Where, number of initial synthetic instance is represented as S, minority and majority class samples counts are denoted as v and u respectively. In second stage, remove the synthetic samples that are closer to the majority class compared to the minority, and the generated instances that are nearer the border created by SMOTE. The ISMOTE step by step process are given as below. Consider the new synthetic instance set as $\hat{y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_S\}$ and $k - th$ attribute value of the $\hat{y}_l, k \in [1, C]$is denoted as $\hat{y}_l^{(k)}$, which are determine the accept or reject the instance created by SMOTE. Let set of majority and minority samples are denoted as $E_p = \{E_{p1}, E_{p2}, \dots, E_{pm}\}$ and $E_q = \{E_{q1}, E_{q2}, \dots, E_{qn}\}$respectively. To decide whether to accept or reject synthetic instance, the distance among the E_{pm} and $\hat{y}_l$ moreover distance among the$E_{qn}$and $\hat{y}_l$ are estimated in initial step using equation (3) and equation (4) respectively.

$$D_n(\hat{y}_l, E_{qn}) = \sum_{k=1}^{C} \sqrt{\left(\hat{y}_l^{(k)} - E_{qn}^{(k)}\right)^2}, n \in [1, v] \quad (3)$$

$$D_m(\hat{y}_l, E_{pm}) = \sum_{k=1}^{C} \sqrt{\left(\hat{y}_l^{(k)} - E_{pm}^{(k)}\right)^2}, \quad m \in [1, u] \quad (4)$$

Through the above expression, two different arrays was created through equation (5) and equation (6).

$$H\left(D_n(\hat{y}_l, E_{q1}), \dots, D_n(\hat{y}_l, E_{qn})\right)_{min} \quad (5)$$

$$H\left(D_m(\hat{y}_l, E_{p1}), \dots, D_m(\hat{y}_l, E_{pm})\right)_{max} \quad (6)$$

Then choose the minimum value off the minority and majority array, if the minority value is small than majority minimum value then synthetic sample is accepted or else it get rejected. That increase the speed of technique and increase performance. In next step, the noise is removed from the accepted synthetic instance$\hat{E} = \{\hat{E}_1, \hat{E}_2, \dots, E_z\}$.The equation (7) computed the distance among $\hat{E}_l$with all real minority E_q.

$$G_n(\hat{E}_l, E_q) = \sum_{j=1}^{v} \sum_{k=1}^{C} \sqrt{\left(\hat{E}_l^{(k)} - E_{qn}^{(k)}\right)^2} \quad (7)$$

Through the above expression the new Tsample set via eliminate the far away samples from the minority is created using equation (8).

$$T = \sum_{i=1}^{n} G_n(\hat{E}_l, E_q) \quad (8)$$

At the same method the distance among the synthetic instance and all majority samples are calculated then remove the minimum distance produced samples for attain high pure data. According to these methods, there are no noisy or borderline samples are augmented in the dataset.

C. Student Performance Prediction using Explainable Deep Learning based Knowledge Distillation Framework

The proposed work introduced the novel EDL_KDF framework for predicting thee student performance. The high accuracy can be attained complex DL model but it utilized more number of parameters thus produced high computational complexity. However, the KD can predicted the student performance with both less computational complexity and high accuracy. The KD produced the light-weight model by reduced the heavy-weight model without decrease the accuracy. Figure 2 shows the proposed framework structure.

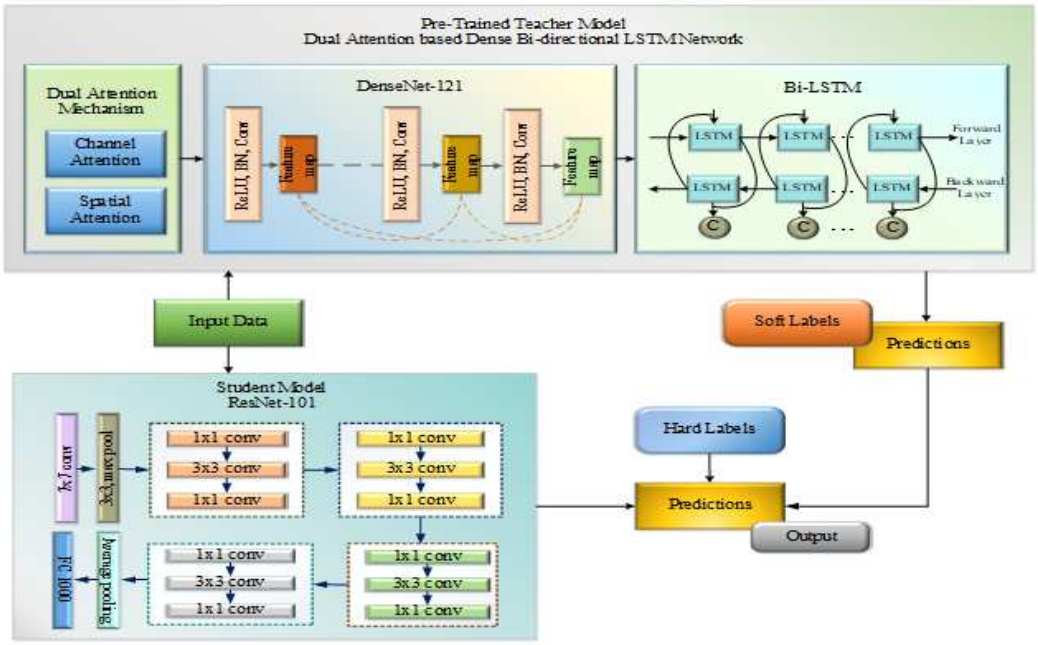


Fig. 2. Structure of EDL_KDF framework

KD employed the pre-trained teacher model (large model) that help to trained student model (small model) by applying the distilled knowledge. For student model training KD have two loss component. First reduces the variance among the predicted and actual labels (hard labels), however the latter reduces the teacher's predictions (soft labels). While distillation process, the probability value is attained using the temperature (T). Initially the logits is extracted from the student model for calculate the probability then it feed in the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

softmax function along with T. The equation (9) expressed the softened softmax function using with temperature.

$$Soft\,max(u_i) = \frac{e(u_i/T)}{\sum_{k=1}^{l} e(u_k/T)} \quad (9)$$

The input data is initially given to the both student and teacher model and receive the logits u_s and u_t. Student and teacher model's probability distribution calculated using equation (10) and equation (11).

$$Pp(T) = Soft\,max(u_t/T) \quad (10)$$

$$Pp(S) = Soft\,max(u_s/T) \quad (11)$$

The probability of teacher and student entropy are control by using the T. The distillation loss is calculated using equation (12).

$$L = \lambda * Lh(Pp(S), O) + (1-\lambda) * Ls(Pp(S), Pp(T)) \quad (12)$$

Where, O, λ, are denoted as actual label (hard label) and weight coefficient respectively. Lh and Ls are computed through the cross entropy. The student learns the appropriate probabilities of specific categories using the pretrained teacher model using an increasing entropy in $Pp(S)$ accordingly with T. The teacher model and student models are mentioned in the below sub-section.

1) Pre-trained Teacher Network using Dual Attention based Dense Bi-directional LSTM Network: DAbDBi-LSTM Network taken as the teacher model that comprises of Dual attention mechanism, DenseNet-121 and Bi-LSTM. The dual attention mechanism integrated the spatial attention and channel attention. The spatial data of feature maps is initially aggregated using average pooling with maximum pooling for channel attention that producing two separate spatial feature maps such as average pooling (C_{Ag}^{S})and max pooling (C_{Mx}^{S}).After that, an integrated network receives this feature map, producing the channel attention $H_S \in \Re^{S\times1\times1}$. The network have hidden layer and multilayer perceptron. By using the reduction ratio the hidden activation size is decreased as$\Re^{S/r\times1\times1}$for minimize the parameter overhead. Element summation is expressed in equation (13) and it employed to create a feature output vector for every feature map and channel attention.

$$H_S(C) = Sigmoid\big(MLP(C_{Ag}^{S}) + MLP(C_{Mx}^{S})\big) \quad (13)$$

The channel information of the feature map is combined using two pooling methods for spatial attention that producing two 2D feature maps such as average pooling ($C_{Ag}^{K} \in \Re^{1\times W\times H}$) and maximum pooling $C_{Mx}^{K} \in \Re^{1\times W\times H}$. Next, spatial attention map is create using equation (14) that joining the two feature maps onto a convolutional layer.

$$H_H(C) = Sigmoid\big(f^{7\times7}\big[C_{Ag}^{S}; C_{Mx}^{S}\big]\big) \quad (14)$$

Where, $f^{7\times7}$ denoted the convolution kernel with 7×7kernel operation. DenseNet-121 architecture contained the dense connection with 121 layers, which make connection among the all-following layers. In order to improve the efficiency of the network's training, each layer obtains significant features that have been learned by previous layers. It utilized less amount of parameter while training and it solve overfitting problem using dense connection. The dense block is the important part of the DenseNet that have Batch Normalization, ReLU and 3×3 Conv that can be expressed using equation (15).

$$D_i = W_i([D_0, D_1, \ldots D_{i-1}]) \quad (15)$$

Where, layers are denoted as $(0,1,\ldots i-1)$ the feature map conjunction generated by DenseNet-121 layers are represented as $[D_0, D_1, \ldots D_{i-1}]$. Operation done over $i-th$ layer is denoted as $W_i(.)$thus, the DenseNet effectively learn the feature map. To realize the deep features produced by the DenseNet-121, Bi-LSTM is introduced in the student model. The LSTM contained three gates such input (S_g), forget (T_g) and output (R_g) gates. For timestep$y$, the LSTM expression are derived in equation (16).

$$\widetilde{D} = tanh\big(K_g V_{xd} + F_{g-1} V_{hd} + Bs_g\big) \quad (16)$$

Where, Bs_g denoted as bias parameter, V_{xd} and V_{hd}represented as weight parameters. In LSTM, $S_g$note that new data get usingD_g. D_{g-1}contained old memory count is addressed by using T_g and by using the point wise multiplication the process is derived in equation (17).

$$D_g = T_g \otimes D_{g-1} + S_g \otimes \widetilde{D}_g \quad (17)$$

Using the output gate, the flow of information among hidden state and memory cell are controlled using equation (19).

$$T_g = R_g \otimes tanh(D_g) \quad (18)$$

The Bi-LSTM process based on the LSTM but it receive the input data in both direction. The Bi-LSTM at the timestep y is expressed in equation (19) and equation (20).

$$\begin{aligned} \vec{T}_y &= \phi\left(K_y V_{xy}^{(f)} + \vec{F}_{y-1} V_{hy}^{(f)} + Bs_y^{(f)}\right) \\ \overleftarrow{T}_y &= \phi\left(K_y V_{xy}^{(b)} + \overleftarrow{F}_{y-1} V_{hy}^{(b)} + Bs_y^{(b)}\right) \end{aligned} \quad (19)$$

By combining the backward and forward LSTM layer, the hidden state is obtain then it passed to the output layer that can be expressed as below.

$$R_g = K_g V_{hq} + Bs_q \quad (20)$$

Where, Bs_q is a bias parameter and V_{hq} is a weight parameters are the output layer's model parameters. Finally the softmax function predicted the student performance. Thus the model effectively learned the student performance data's presented in the dataset through the DAbDBi-LSTM Network.

2) Prediction make using ResNet-101 Student Network: Resnet-101 has an ability to optimize the residuals across input and targeted convolution features sets that contained

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

101 layers. So that target features are effectively attained effectively compare than other network. ResNet-101 greatly decrease the deeper network's parameter by applying the residual optimization, which also decrease the amount of layers in the network. Data that is not learned in the preceding layer of the ResNet architecture is applied from the old to its new layer through the ResBlock layer. The blocks of the Resnet architecture is ResBlock layer, which provide residual values into the subsequent layer. A two-layer activation includes a skip that provides value across the weight layers as well as the Relu activation code. This skip affects the system account. The equation (21) shows the nonlinear Relu function.

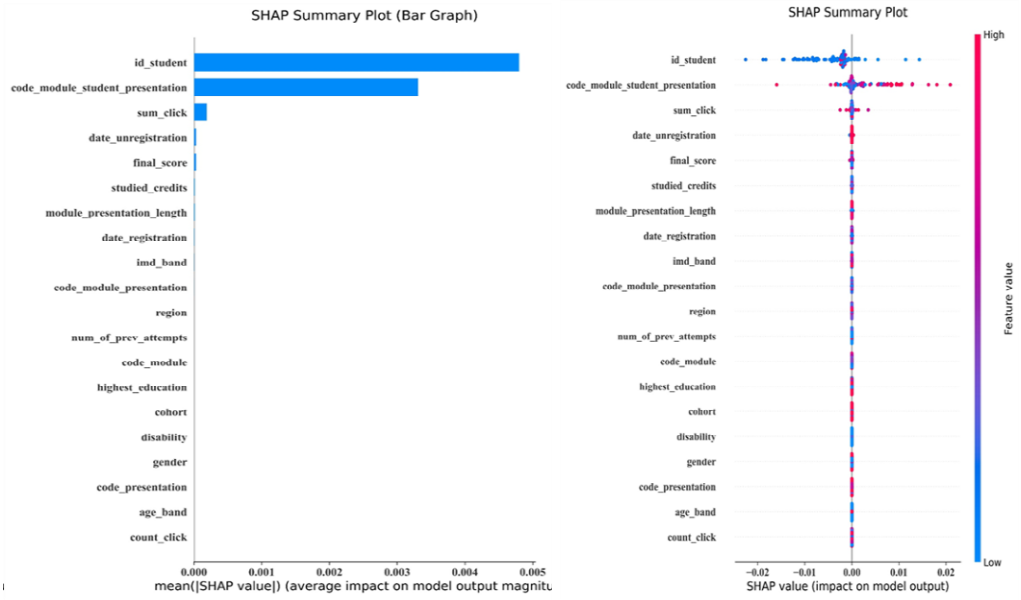
$$R = w_{2\sigma}(w_1 u) \tag{21}$$

Output is attained through equation (22) that concatenate the second Relu value.

$$O = R(u, \{w_i\}) + u \tag{22}$$

Where, u and O are represented the input and output vector respectively. At the end the softmax activation function predicted the student performance by using the ResNet-101 with the help of DAbDBi-LSTM teacher network.

3) SHapley Additive Explanations: SHAP is generally used for analyze the feature major contribution for the prediction output. A heatmap illustrates that how feature values are impact the prediction outcome. Figure 3 shows the model's features mean absolute Shapley value and magnitude outcome.



(a) (b)

Fig. 3. Analysis of (a) Mean absolute and (b) magnitude outcome of SHAP value

From the SHAP analysis, the feature id_student is observed that feature magnitude predicts the performance of the teams at approximately 0.005. Proposed prediction model received the positive and negative influence due to the low feature value from -0.02 till 0.014 for id_student feature. At the same time low feature value of code_module_student_presentation feature influence positive prediction from -0.006 to 0.02. This SHAP supports to determining the corresponding importance of several features as well as how these individuals affect the output of the EDL_KDF prediction model.

IV. Result And Discussion

In this section, the performance of the EDL_KDF framework is analyzed for proving the efficiency of the proposed work compared to other existing student performance prediction framework. The proposed EDL_KDF framework implemented through using the PYTHON tool.

A. Dataset description

Open University Learning Analytics Dataset (OULAD), which is a standardizing and benchmarking dataset. The dataset contained 32,593 students log data's like demographics, interaction data, registrations and assignments. The dataset is download using https://www.kaggle.com/datasets/anlgrbz/student-demographics-online-education -dataoulad link. The Open University consider social sciences and science, technology, engineering, and mathematics (STEM) courses for gathered the student activity. The student activity data's are recorded from 2013 to2014.

B. Analysis of result and comparison

The performance of the EDL_KDF framework is analyzed via several DL performance metrics such as accuracy, precision, recall, F1-Score and Receiver Operating Characteristic (ROC) Curve, these analysis graphs are given in Figure 4.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

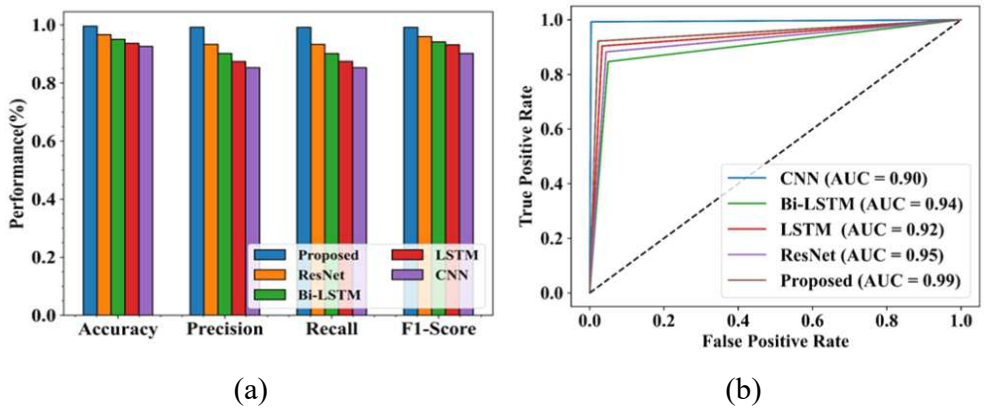


Fig. 4. Analysis of EDL_KDF framework using (a) various performance metrics (b) ROC curve

The DL performance metrics are commonly used to measure the effectiveness, robustness and efficiency of the prediction model. The high results in terms of F1-score, precision, recall and accuracy produced DL model outperform in the prediction. The EDL_KDF framework attained 99.6% of accuracy, 99.21% of precision, recall and F1-score. At the same time, the framework attained 0.99 Area under Curve (AUC) in ROC analysis. The obtained performance are shows better outcome compared to other DL model. Figure 5 shows the accuracy and loss curve analysis.

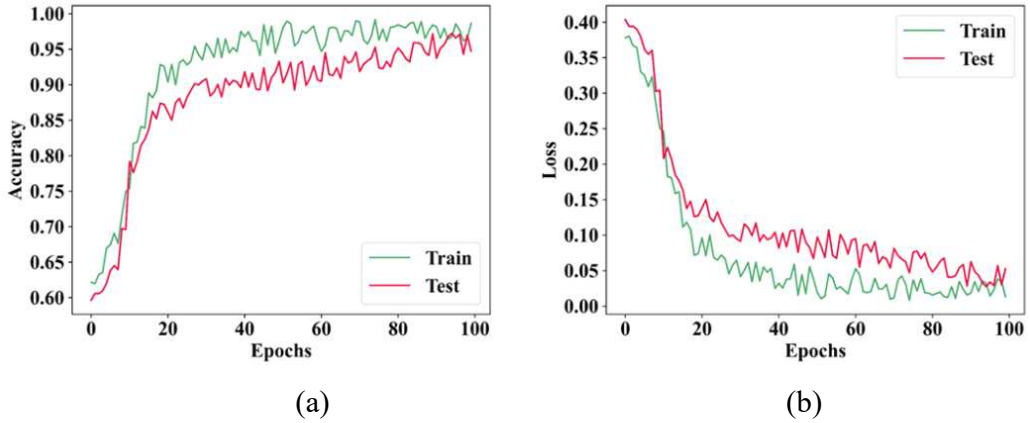


Fig. 5. Analysis of (a) Accuracy curve and (b) Loss curve

The EDL_KDF framework accuracy and loss are analyzed based on the 100 epochs in terms of both training and testing. After attaining the 40 epochs, the accuracy of the prediction is gradually increased as well as loss of the prediction model majorly decreased. Thus help to predict the student performance over the OULAD dataset. Figure 6 shows the analyzed the proposed framework performance regarding error metrics.

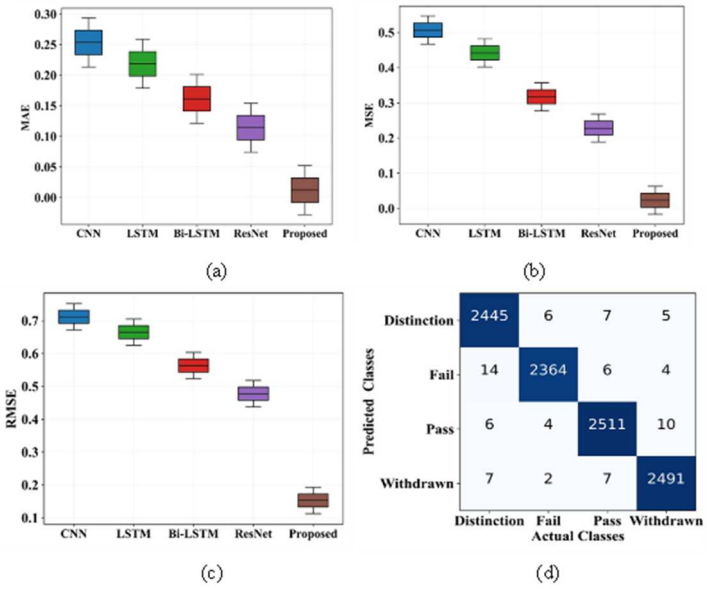


Fig. 6. Comparison and performance analysis based on (a) MAE (b) MSE (c) RMSE and (d) Confusion matrix

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

The MSE, MAE and RMSE are generally utilized for examine the error rate of the prediction model. The obtained value of the error metrics very low while the prediction model attained better performance. The EDL_KDF framework produced less error value such as 0.023% of MSE, 0.0122% of MAE and 0.1528% of RMSE. EDL_KDF framework accurately predicted into its corresponding classes. Less amount of prediction has been wrongly predicted into the other classes. From overall performance analysis, the EDL_KDF framework give the better performance over the student performance prediction.

C. Discussion

The performance of the EDL_KDF framework is discussed in this section. The existing student performance prediction model issues are effectively reduced by the proposed work. The overfitting problem [13, 16] was reduced by the DenseNet-121, which is utilized in the student model of the EDL_KDF framework. Dataset class imbalance problem [14] is overcome by the ISMOTE. Accuracy and computational complexity issues [15, 17] are solved by the novel EDL_KDF framework. The EDL_KDF framework make a prediction using the ResNet-101 network with the distillation knowledge of DAbDBi-LSTM network, thus effectively decrease the computational complexity without decreasing the prediction accuracy. Table 1 contained the comparison analysis with existing prediction model.

Through the analysis, proposed EDL_KDF framework effectively predicted the student performance with high accuracy, precision and recall with high computational complexity. Thus, the proposed work outperform compared to other existing prediction model.

TABLE I. COMPARISON ANALYSIS WITH EXISTING STUDENT PREDICTION MODEL

Author and References	Model	Accuracy (%)	Precision (%)	Recall (%)
Waheed et al. [9]	Deep ANN	84-93	93-96	69-86
Giannakas et al. [10]	DNN	82.39	77.08	68.52
Asselman et al. [11]	Adaboost, XGBoost and random forest ensembled algorithm	71.40	73.97	78.15
Rodríguez-Hernández et al. [12]	ANN	82	-	-
Kusumawardani et al. [13]	DL based transformer encoder	83.17	-	-
Kukkar et al. [14]	RNN + LSTM + RF	97	-	-
Almahdi et al. [15]	V. DL based an Optimized Predictive Academic Performance Approach	98.94	-	-
Proposed	VI. EDL_KDF framework	99.6	99.21	99.21
	VII.			

V. CONCLUSION

In this paper, the proposed EDL_KDF framework effectively predicted the student performance over the OULAD dataset. The EDL_KDF framework initially increased the quality of the dataset at the pre-processing stage, thus improved the prediction accuracy. After that, novel EDL_KDF framework attained the accurate outcome in the student performance prediction through the KD model. The better accuracy was attained by the EDL_KDF framework with the help of both DAbDBi-LSTM network and ResNet-101. The DAbDBi-LSTM network was pre-trained in the EDL_KDF framework, which give the distillation knowledge to the ResNet-101 student model for predicting the output. Through the process, the less complex ResNet-101 model extracted the knowledge of complex DAbDBi-LSTM network, so that the computation

Complexity of the proposed student performance prediction was increased as well as it attain high accuracy in prediction. Finally the SHAP was introduced for analyses the every features in model that help to know the impact of the particular feature in the prediction. The EDL_KDF framework evaluated via several performance metrics and it achieved 99.6% and 99.21% of accuracy and precision over the OULAD dataset respectively. In future, the generalization loss of the KD model is focus the improvement student performance prediction performance for unknown data.

REFERENCES

[1] R. Bhuvana, & P. S. Aithal, "Impact of COVID-19 on Redefining the Services of Educational Institutions using Ubiquitous Technology, International Journal of Management," Technology and Social Sciences (IJMTS), vol. 5, no. 2, pp. 266-282, 2020.

[2] S. A. Aljawarneh, "Reviewing and exploring innovative ubiquitous learning tools in higher education", Journal of computing in higher education, vol. 32, pp. 57-73, 2020.

[3] A. Kumar, R. Krishnamurthi, S. Bhatia, K. Kaushik, N. J. Ahuja, A. Nayyar, & M. Masud, "Blended learning tools and practices: A comprehensive analysis," Ieee Access, vol. 9, pp. 85151-85197, 2021.

[4] S. Butt, A. Mahmood, S. Saleem, S. A. Murtaza, S. Hassan, & E. Molnár, "The Contribution of Learner Characteristics and Perceived Learning to Students' Satisfaction and Academic Performance during COVID-19," Sustainability, vol. 15, no. 2, p. 1348, 2023.

[5] R. A. Rasheed, A. Kamsin, & N. A. Abdullah, "Challenges in the online component of blended learning: A systematic review," Computers & Education, vol. 144, p. 103701, 2020.

[6] P. G. Altbach, L. Reisberg, & L. E. Rumbley, Trends in global higher education: Tracking an academic revolution. Brill, vol. 22, 2019.

[7] C. Lim, H. Ab Jalil, A. Ma'rof, & W. Saad, "Peer learning, self-regulated learning and academic achievement in blended learning courses: A structural equation modeling approach," International Journal of Emerging Technologies in Learning (IJET), vol. 15, no. 3, pp. 110-125, 2020.

[8] B. Sekeroglu, K. Dimililer, & K. Tuncal, "Student performance prediction and classification using machine learning algorithms," In Proceedings of the 2019 8th International Conference on Educational and Information Technology, pp. 7-11, 2019.

[9] H. Waheed, S. U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, & R. Nawaz, 'Predicting academic performance of students from VLE big data using deep learning models," Computers in Human behavior, vol. 104, p. 106189, 2020.

[10] F. Giannakas, C. Troussas, I. Voyiatzis, & C. Sgouropoulou, "A deep learning classification framework for early prediction of team-based academic performance," Applied Soft Computing, vol. 106, pp. 107355, 2021.

[11] A. Asselman, M. Khaldi, & S. Aammou, "Enhancing the prediction of student performance based on the machine learning XGBoost algorithm," Interactive Learning Environments, vol. 31, no. 6, pp. 3360-3379, 2023.

[12] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, & E. Cascallar, "Artificial neural networks in academic performance prediction:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

Systematic implementation and predictor evaluation," Computers and Education: Artificial Intelligence, vol. 2, p. 100018, 2021.

[13] S. S. Kusumawardani, & S. A. I. Alfarozi, "Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities," IEEE Access, vol. 11, pp. 18960-18971, 2023.

[14] A. Kukkar, R. Mohana, A. Sharma, & A. Nayyar, "A novel methodology using RNN+ LSTM+ ML for predicting student's academic performance," Education and Information Technologies, pp. 1-37, 2024.

[15] A. A. Almahdi, & B. T. Sharef, "Deep Learning Based An Optimized Predictive Academic Performance Approach," In 2023 International Conference on IT Innovation and Knowledge Discovery (ITIKD), pp. 1-6, 2023. IEEE.

[16] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, & E. Cascallar, "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation," Computers and Education: Artificial Intelligence, vol. 2, p. 100018, 2021.

[17] S. S. Kusumawardani, & S. A. I. Alfarozi, "Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities," IEEE Access, vol. 11, pp. 18960-18971, 2023.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 06:56:58 UTC from IEEE Xplore. Restrictions apply.

MĀPAN-Journal of Metrology Society of India
https://doi.org/10.1007/s12647-024-00758-x

ORIGINAL PAPER

Size Optimization of Grid-Tied Hybrid Energy System by Employing Forecasted Meteorological Data

P. Anand[1], B. Sharma[2,3]* and M. Rizwan[2]

[1]Department of Electronics and Communication Engineering, Bhagat Phool Singh Mahila Vishwavidyalaya, Khanpur Kalan, Sonipat 131305, India

[2]Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India

[3]Department of Electrical and Electronics Engineering, Krishna Institute of Engineering and Technology, Ghaziabad 201003, India

Received: 27 February 2024 / Accepted: 22 May 2024

© Metrology Society of India 2024

Abstract: Embracing hybrid energy systems (HES) to ensure access to clean, reliable, and cost-effective energy is necessary for nations that are striving for sustainable development. By leveraging precise meteorological data from forecasts, the HES can be rendered more accurate. Thus, firstly, the research presented here employed four machine learning approaches, such as Gaussian process regression (GPR), support vector regression, extreme gradient boosting, and decision trees, to carry out hourly forecasting of meteorological data over a year. The results obtained revealed that the GPR outperformed the other three forecasting models. For this reason, the forecasted meteorological data acquired from GPR is employed in the sizing of the HES. Tunicate swarm algorithm (TSA), a recently developed method, is applied to perform the size optimization of HES capable of meeting the energy necessities at remote sites in the Indian province of Uttar Pradesh. Following a comparative study of TSA, particle swarm optimization, and harmony search, TSA proved to yield a better outcome. Additionally, the simulation result showed a 0.33% cut in the per-unit cost of energy when forecasted data becomes the basis for the optimization of system size.

Keywords: Hybrid energy system; Forecasting; Tunicate swarm algorithm; Machine learning approaches

1. Introduction

The rise in greenhouse gases (GHG), involving methane, carbon dioxide, carbon monoxide, nitrous oxide, fluorinated gases, and others, has sparked concerns around the globe because it is directly indicative of climate change. The main root cause of GHG emissions, resulting in catastrophic consequences such as extreme weather, temperature rises, and ecological disruption, has been identified as the result of the burning of fossil fuels. The surge in energy consumption worldwide is also attributed to factors like urbanization, economic expansion, population growth, and a greater reliance on appliances and technology. To cope with these issues, efficient and environmentally conscious energy solutions ought to be devised with a view to complying with growing demand while limiting adverse environmental effects. The deployment of renewable energy is a crucial tactic to combat cutting-edge issues while paving the path for a resilient and cleaner future [1–4]. However, due to environmental conditions featuring fluctuating wind speed and sun irradiation, leaning solely on renewable energy sources ($R_E S$) seems to be less effective than conventional energy sources and grid-tied systems. The aforementioned factors may cause outages or cutbacks in generated energy at different instances of day. Hybrid energy system (HES) has been pitched as a pragmatic answer to this problem. HES blend a multitude $R_E S$ with energy storage or backup devices to improve reliability [5–7].

An essential aspect of the design of HES entails figuring out the optimal system size for a particular load demand. Researchers have come up with an assortment of optimization tactics, covering deterministic methods, metaheuristic algorithms, and software tools. For sizing HES, popular and effective software programs like hybrid

*Corresponding author, E-mail: bandana.elec@gmail.com

Published online: 06 July 2024

optimization model for multiple energy resources (HOMER), hybrid optimization by genetic algorithm (HOGA), RETscreen, and TRYNS, are often utilized. Specifically, HOMER is a potent tool for developing HES systems, finding the optimal size, and conducting economic evaluations as well [8, 9]. One of the primary downsides of the HOMER software is that users are unable to modify the modeling equation, sizing procedure, algorithms, or calculations. On the other hand, deterministic methods like graphical, probabilistic, analytical, iterative, and linear programming are more valuable than software tools since they have been applied in an extensive set of HES research. Conversely, deterministic approaches are limited to selecting a local optimal solution; they are incapable of stumping up a global optimal solution [10, 11]. In recent times, many meta-heuristic methods prompted by natural behavior have been applied with the aim of optimally sizing HES. A handful of these are the following ones: biogeography-based optimization (BBO), particle swarm optimization (PSO), grey wolf optimization (GWO), artificial bee colony (ABC), artificial bee swarm optimization (ABSO), simulated annealing (SA), ant colony optimization (ACO), harmony search (HS), bacterial foraging algorithm (BFA), mine blast algorithm (MBA), and others [12–21]. These algorithms can discover optimum solutions more precisely than deterministic techniques, avoiding the trapping of local optima [20]. Significant research has been done on the most effective ways to size HES, utilizing a range of optimization strategies. An overview of a handful of these investigations is provided in the subsequent paragraphs.

The most optimal HES of solar photovoltaic (S_{PV}), wind, diesel, and battery storage systems were found by employing the grasshopper optimization method [22]. The costs were maintained to a minimum by limiting the cost of energy (CE) while considering the probability of a power outage into consideration. The entire cost of HES, encompassing of a storage system, S_{PV} panels, diesel generator (DG), and wind turbine (WT), has been drove down with the aid of the PSO technique [23]. The whale optimization algorithm (WOA) and PSO approaches were applied in the design of a hybrid microgrid that included S_{PV}, fuel cells (FC), and biowaste [24]. The entire net present cost (NPC) was the fitness function in this design problem involving constraints relating to availability and loss of power supply probability (LPSP). A HES with a DG, WT, S_{PV}, and battery was suggested for the Dakhla area of Morocco. The goal of the research is to reduce investment expenses by utilizing the Equilibrium Optimizer (EO) while keeping into account multiple constraints such as renewable fraction, availability, and reliability. The computational outcomes were contrasted with those of optimization methods derived from Sooty Terns, GWO, Artificial Electric Fields (AEF), and harris hawks optimizer (HHO) [25]. In another instance, A HES for a microgrid in Egypt's Farafra region was developed using the improved archimedes optimization algorithm (IAOA). Slashing the NPC while attaining three critical constraints—renewable energy fraction, power supply reliability, and system availability—was the main objective. The outcomes of the simulation were contrasted with the outcomes of popular algorithms, especially, original archimedes optimization algorithm (AOA), AEF, EO, GWO, and hydrogen hydroxide optimizer. The results exhibit how effectively the IAOA copes with the design problem while showing how efficient it is in contrast to other algorithms [26].

A comprehensive assessment of HES with large-scale battery storage to enhance a smart city's sustainability was carried out by Nuvvula et al. Their work optimized the sizing of several R_ES within the system using mutation-enabled quadratic particle swarm optimization (QPSO). Limiting power supply losses, optimizing the net present value (NPV), and lowering energy expenses were the main goals [27]. A variation of the dragonfly algorithm termed the modified dragonfly algorithm (MDA) was employed in mathematical modeling with the intention of reducing the NPV of power systems over the course of their operational life. After weighing multiple possibilities, the model concluded that the best cost-effective way for bolstering dependability is to use the HES with grid interaction [28]. A methodology for the deterministic and probabilistic design of a S_{PV} module, biomass gasifier (B_M), and battery HES was proposed by Kumar and colleagues in a different study. In an effort to cut energy expenditures, they applied three optimization algorithms: the JAYA method, the aggrandized class topper optimization (ACTO), and the PSO [29]. Diab et al. [30] presented a simulation model for a HES microgrid along with battery storage. Their goal was to hone the HES's efficiency, dependability, and cost-effectiveness while figuring out the optimal system size. They employed revolutionary optimization methods such as the moth flame optimization (MFO), water cycle algorithm (WCA), and WOA. The simulation results showed that WOA performed more proficiently than the other algorithms. Tabak et al. [31] undertook research to optimize a HES comprising a B_M, WT, and S_{PV} module, components. They scaled the system to suit the energy needs of the faculty building by optimizing it by means of a genetic algorithm (GA). Unlike earlier research, they put more emphasis on reliability and optimized the system size by adopting GWO. They noticed that the GWO approach performed superiorly to the GA outcomes.

Divya et al. [32] used the black widow optimization (BWO) algorithm to optimize HES under uncertainty constraints. With rigorous modeling and realistic validation, it emphasizes combining S_{PV}, wind, and hydro

sources while retaining a balance between cost-effectiveness and reliability. BWO showcases its potential for sustainable energy solutions by optimizing fitness to cut annual operating costs while effectively complying with specified loads using MATLAB. Gupta et al. [33] proposed a novel spoonbill swarm optimizer (SSO) to identify the optimal hybrid renewable-energy microgrid (MG) system for an Indian community. Considering technological, financial, social, environmental, and reliability considerations, the selected model comprises S_{PV}, wind, B_M, battery, DG, and electric vehicle (EV) components. Results corroborate the optimized MG system's efficiency across multiple criteria, demonstrating the SSO's superiority over competing methods in reducing convergence time and iterations. With the intent of maximizing reliability, minimizing expenses, and maintaining a sufficient power supply, a hybrid approach termed light spectrum optimizer (LSO) and deep attention dilated residual convolutional neural network (DADRCNN), notably LSO-DADRCNN, is applied by Ramesh et al. [34] for optimizing the sizing of HES with EV. By employing DADRCNN for EV energy demand forecast and LSO for system sizing, the approach trumps previous approaches such as heap-based optimizer (HBO), wild horse optimizer (WHO), and PSO in the realm of accuracy and cost reduction, attaining 96% efficiency and saving $1.7. Messini et al. [35] examined S_{PV}-hydrogen-based HES integration in oil and gas sites, emphasizing efficiency and carbon reduction. By optimizing with Fick's law algorithm (FLA), it cuts down costs, eradicates power loss, and lowers carbon dioxide emissions by 80%, promoting S_{PV} -hydrogen-based HES for sustainable energy policies. In Kaele, Cameroon, Koholé et al. [36] investigated an off-grid HES involving S_{PV}, wind, batteries, and DG for household, multimedia, and healthcare centers. Four meta-heuristic algorithms, such as charged system search (CSS), colliding bodies' optimization (CBO), water evaporation optimization (WEO), and teaching–learning-based optimization (TLBO), were applied to optimize HES in view of NPC. Results revealed that TLBO outperforms the other approaches.

A major impediment to the design of HES is the problem of volatility in R_ES, which has led to the exploration of several solutions. A bibliographical evaluation was carried out by Lei et al., [37] who explored recent findings and breakthroughs in wind speed (WS) and power forecasts. A novel hybrid electrical engineering method for short-term wind forecasting via wavelet transform is presented by Catalao et al. [38]. The WS and GHI statistics from the past year often serve as the cornerstones of the optimization strategies outlined in the literature. It is acknowledged, therefore, that historic meteorological data diverge significantly from projections made now. With the goal of improving energy efficiency, a number of models involving uncertainties have been devised. In this vein, Hocaoglu et al. examined the influence of global horizontal irradiance (GHI) forecasting on optimal system design efficiency, whereas Tascikarao glu et al. and Khalid et al. [39–41] used wind power prediction for the development of HES. Zhang et al. [42] proposed a novel methodology that combines HS, SA, and chaotic search to figure out the ideal size of grid-connected HES. To augment the precision of the size-optimized design, weather and load forecasts were executed leveraging artificial neural networks. To forecast long-term weather conditions, Alfonso et al. [43] offer a sizing methodology that makes use of long-term memory cells. Maleki et al. [44] reported a study testing a HES employing iterative neural networks to predict meteorological data for fulfilling household water needs. The study involved HS and a mix of HS and chaotic search. Gupta et al. [45] suggested solar-wind forecasting using artificial neural networks and size optimization using the BBO technique. Using projected data instead of actual data, they emphasized the benefits of HES size optimization.

An extensive literature review underlines how crucial it is to thoroughly evaluate external, technical, and financial energy sources while setting up HES. Recently, intelligent techniques have gained a growing following due to their stunning outcomes. Much more research needs to focus on grid-tied HES adopting intelligent techniques, although simulation tools like the HOMER are frequently applied in studies dwelling on grid-connected scenarios. Researchers have paid scant attention to the generation of power from biogas and biomass, in spite of their immense potential—especially in rural zones. Further, it is pointed out that multiple investigations have demonstrated that intelligent techniques may solve optimization problems and often use data from a single day that has not undergone an intensive statistical investigation. Therefore, more specialized work in HES needs to be done to fill this research gap. Moreover, a lot of the size optimization methods that are currently in the literature rely on historical data for solar irradiance and wind speed. On the other hand, historical meteorological data shows notable differences from present values. It has also been noticed that previously published studies have rarely used machine learning (ML) techniques for forecasting solar, wind, and temperature data to size HES. More specialized work in HES needs to be done to fill this research gap. A significant research gap exists because previously published studies have rarely used ML techniques for forecasting solar, wind, and temperature data to size HES.

In order to address the aforementioned problems, the current study first uses ML techniques, specifically Gaussian process regression (GPR), extreme gradient boosting (XGBT), support vector regression (SVR), and decision tree (DT), to perform hourly forecasts of temperature and

GHI of selected site for a year. The most appropriate solution is subsequently determined. On top of that, the optimal HES has been developed using the predicted GHI and temperature data. Employing a recently developed method known as Tunicate swarm algorithm (TSA), the optimal sizing of HES is carried out both with and without the predicted data. Furthermore, a comparison of the TSA, HS, and PSO techniques has been done and identified the appropriate outcome. More precisely, the key findings of the present work are encapsulated as follows:

The research work addressed the problem of optimal component sizing in HES by utilizing the recently established TSA algorithm. An assessment of TSA's effectiveness has also been undertaken by comparing its outcomes with those of well-known approaches such as HS and PSO.

Researchers usually disregarded biogas and biomass-based power generation, despite understanding that they indicate substantial and likely advantageous resources, most notably in rural regions. By incorporating biogas and biomass generators into the enacted HES in tandem with the scheduling of their operation, this study bridges this research gap.

The optimization of HES component sizing is carried out through the use of forecasted data, an idiosyncratic facet rarely found in existing literature. This distinctive approach enhances the effectiveness of size optimization for HES components. To facilitate this, assorted ML algorithms have been deployed to forecast temperature and sun irradiation. On top of that, the outcomes obtained by forecasted and historical data are contrasted, and the most appropriate solution is found.

2. Tunicate Swarm Algorithm

The tunicate colonies—a species of marine animal that inhabits piers, rocks, or the underneath of ships and boats—serve as a metaphor for the TSA algorithm. Tunicates are discernible from a distance owing to their bright blue-green bioluminescence, which usually mimics tiny, colorful blobs. These creatures have two ends: an open end and a closed end. The open end generates jet propulsion by using atrial siphons. Tunicates boost themselves vertically in the water by putting out bursts of fluid along with an explosive force that is gauged in millimeters. Even though they are so small, tunicates have an amazing knack for detecting food sources in the water, regardless of whether they are cognizant of them. Tunicates' primary functions embrace jet propulsion and swarm intelligence, which facilitate their ability to figure out the best places to find food [46].

2.1. Mathematical Modeling of TSA Technique

The mathematical modeling of the initial course of action of the tunicates, specifically their jet propulsion, must satisfy three requirements. These standards are meant to avert disputes between tunicates, modify potential tunicate positions, and render closeness to the intended tunicate. However, in order to seek out the optimum and best-suited solution, swarm behavior entails keeping an eye on the existence of fellow seekers. The sections that proceed delve into more detail about how these behaviors are illustrated mathematically [46].

Step 1: Avoiding Search Agent Conflicts

To avert the conflicts among search agents (S_A), vector $\overrightarrow{M}$ is put to use in used to determine the new location of S_A.

$$\overrightarrow{M} = \frac{\overrightarrow{F_G}}{\overrightarrow{F_S}} \tag{1}$$

$$\overrightarrow{F_G} = R_2 + R_3 - \overrightarrow{A} \tag{2}$$

$$\overrightarrow{A} = 2.R_1 \tag{3}$$

where, $\overrightarrow{F}_G$ defines the gravity force; $\overrightarrow{A}$ signifies deep ocean advection;R_1, R_2, R_3 are indicative of the random values ranging from 0 to 1; The social forces among S_A are encapsulated by $\overrightarrow{F_S}$.

The subsequent equation has been applied for calculation of the vector $\overrightarrow{F_S}$:

$$\overrightarrow{F_S} = [N_{mn} + R_1(N_{mx} - N_{mn})] \tag{4}$$

where;N_{mn}, N_{mx} signifies initial and subservient speeds required for social interaction.

Step 2: Moving Ahead in the Path of Best Neighbor

After abstaining from neighbor contentions, the S_A moves ahead in following the path of best neighbor.

$$\overrightarrow{d} = | \overrightarrow{P_S} - R \cdot \overrightarrow{P_T}(i) | \tag{5}$$

where, $\overrightarrow{d}$ indicates the distance between food source and S_A; i is the present iteration. The position of optimal food source is encapsulated by $\overrightarrow{P_S}$; the tunicate position is denoted by $\overrightarrow{P_T}(i)$; and R signifies a random value in the range of 0 and 1.

Step 3: Converge towards the Best S_A.

e_A may preserve its position with regard to its optimal location, which happens to be close to a food source.

$$\overrightarrow{U_p}(i) = \overrightarrow{P_S} + \overrightarrow{M} \cdot \overrightarrow{d}, \quad \text{if} \quad R \geq 0.5 \tag{6}$$

$$\overrightarrow{U_p}(i) = \overrightarrow{P_S} - \overrightarrow{M} \cdot \overrightarrow{d}, \quad \text{if} \quad R < 0.5 \tag{7}$$

where; $\overrightarrow{U_p}(i)$ symbolizes the tunicate's updated position relative to the food source position $\overrightarrow{P_S}$.

Step 4: Behavior of Swarm

In order to mimic the swarm behavior of tunicates mathematically, the positions of the leftover S_A are updated in accordance with the optimal location of the first of the two optimal solutions, displacing the other S_A's position unalterable. The formula listed below can be applied to demonstrate the tunicate swarm behavior:

$$U_P\left(\overrightarrow{i}+1\right)=\frac{\overrightarrow{U_P}(i)+\overrightarrow{U_P}(i+1)}{(2+R_1)} \tag{8}$$

After satisfying the predetermined stopping criteria, the TSA algorithm successfully converges to the optimal design for the given problem as the swarm behavior iteratively modifies the positions of other S_A toward the best optimal solution. Because of its collective behavior, which enables adaptive and efficient adjustments to changes, TSA exhibits remarkable resilience in dealing with dynamic and uncertain environments. Further, the steps of the TSA algorithm are outlined below.

- **Step 1:** Initialization of tunicate population $\overrightarrow{U_P}$.
- **Step 2:** Select the maximum number of iterations as well as the initial parameters.
- **Step 3:** Compute the fitness value for each S_A.
- **Step 4:** Upon calculating the fitness value, the optimal search agent within the specified search space is investigated.
- **Step 5:** Update each S_A's position by applying Eq. 8.
- **Step 6:**In a specific search space, adjust the updated S_A that crosses the boundary.
- **Step 7:** Determine the current fitness value of the search agent. Update U_P if there is a better solution than the prior best option.
- **Step 8:** The algorithm comes to an end if the stopping criterion is met. If not, go back and do steps 5 through 8.
- **Step 9:** Provide the most optimal solution that has been found thus far.

3. Modeling of HES Components

The primary emphasis of this research is on development of the most optimal viable HES that integrates B_M, biogas generator (B_G), and S_{PV} energy sources. Formulating mathematical expressions for each component is essential to efficiently designing and figuring out the optimal size for the HES. Consequently, these mathematical expressions are given a thorough overview in the sections that follow.

3.1. S_{PV} Panels

A S_{PV} system's output power ($P_{SP_V}(t)$) can be affected by several factors, such as the panel's size, the amount of sunlight it receives each day, the cell temperature, and its placement. This relationship can be expressed using Eq. 9 [47]:

$$P_{SP_V}(t)=e_{SP_V}\times A_{SP_V}\times R_S(t)\ (1-0.005(T_{CL}-25)) \tag{9}$$

where the efficiency of the S_{PV} modules is encapsulated by e_{SP_V}, while A_{SP_V} symbolizes the surface area of these modules. Additionally, the solar radiation received by the S_{PV} panel is represented by $R_S(t)$, and T_{CL} stands for the cell temperature. The mathematical expression for computing the cell temperature is mentioned below:

$$T_{CL}=T_{AT}+T_{CO}-200.8*R_S(t) \tag{10}$$

where, T_{AT} and T_{CO} stands for the ambient temperature and nominal operating cell temperature are represented by respectively.

3.2. BM

The power output of the B_M at a given hour "t" symbolizes as "P_M (t)" is calculated using the following equation [48]:

$$P_M=\frac{c_M\times d_M\times e_M\times 1000}{365\times 860\times h_M} \tag{11}$$

Biomass's calorific value is encapsulated by "c_M", "d_M" represents the quantity of available biomass, "e_M" symbolizes the efficiency of the biomass conversion process, and "h_M" stands for the number of hours the B_M operates daily.

3.3. BG

The power output of the B_G at a specific time "t" denoted as "P_G (t)" is computed using the following formula [49]:

$$P_G=\frac{c_G\times d_G\times e_G}{h_G\times 860} \tag{12}$$

Here, "c_G" symbolizes for calorific value of biogas, the daily availability of biogas is encapsulated by "d_G"; "e_G" stands for the overall efficiency of the biogas-to-electricity conversion process, and the daily operational hours of the B_G is encapsulated by "h_G".

3.4. Battery

The method of operation for the battery can be charging or discharging, contingent upon the balance between energy production and consumption. When the hourly electrical

load requirement exceeds the amount of electricity generated by the $R_E S$, the excess energy is stored in the battery during the charging mode. On the other hand, in the discharging mode, the battery releases stored energy to fulfil the load demand when it surpasses the amount of electricity supplied by $R_E S$. Depending on whether the battery is in the charging or discharging condition, Eqs. 13 to 15 calculate the battery's capacity at hour "t".

In the state of charging, the energy stored in the battery is indicated in equation as follows:

$$E^{s}_{Bt}(t) = E^{s}_{Bt}(t-1) + \left[E^{EE}_{B_M}(t) + E^{EE}_{B_G}(t) + E^{EE}_{SP_V}(t)\right] \times e_C \tag{13}$$

In the above equation, the energy stored in the battery at hour "t" is encapsulated by "$E^{s}_{Bt}(t)$". The terms "$E^{EE}_{B_M}(t)$," "$E^{EE}_{B_G}(t)$" and "$E^{EE}_{SP_V}(t)$" define the surplus energy generated by B_M, B_G, and S_{PV} systems, respectively, after meeting the electrical load demand. The battery charging efficiency is encapsulated by "e_C".

$$E^{s}_{Bt}(t) = (1 - \cap) \times E^{s}_{Bt}(t-1) - E_{OD}(t) \tag{14}$$

$$E_{OD}(t) = \frac{E_{DE}(t)}{e_i \times e_D} \tag{15}$$

Here, '$\cap$' denotes the rate at which the battery self-discharges at a given hour 't.' '$E_{OD}(t)$' stands for the overall electrical energy demand. 'e_D' and 'e_i' symbolizes for the efficiency of discharging the battery and the efficiency of the inverter, respectively. The energy demand that is not fulfilled by $R_E S$ is encapsulated by '$E_{DE}(t)$'.

$$E_{DE}(t) = E_{ED} - [E_{B_M}(t) + E_{B_G}(t) + (E_{SP_V}(t) \times e_i)] \tag{16}$$

'E_{ED}' denotes the hourly electrical load demand. $E_{SP_V}(t)$, $E_{B_M}(t)$, and $E_{B_G}(t)$ denotes energy generated by $S_{PV}, B_M,$ and B_G system, respectively.

3.5. Utility Grid

In case of grid-tied HES, the grid can operates in two distinct ways. In the initial mode, it has the capability to supply more electricity to the HES if the combined output from $R_E S$ and the battery is not enough to cope with the energy demand. Equation (17) provides a mathematical expression for this scenario [50, 51].

$$E_{GP}(t) = E_{ED}(t) - [E_{B_M}(t) + E_{B_G}(t) + ((E_{SP_V}(t) + (E^{d}_{Bt}(t) - E_{Btn})) \times e_i)] \tag{17}$$

The grid-supplied energy is encapsulated by "$E_{GP}(t)$". "E_{Btn}" symbolizes the minimum required battery storage capacity.

The second mode involves returning the excess energy generates by HES to the utility grid, which is determined by the following equation:

$$E_{GS}(t) = [E_{B_M}(t) + E_{B_G}(t) + (E_{SP_V}(t) + (E_{Btx} - E^{d}_{Bt}(t)) \times e_i] - E_{ED}(t) \tag{18}$$

"$E_{GS}(t)$)" reflects the excess energy that must be returned to the grid. The battery storage system's maximum capacity is indicated by the symbol "E_{Btx}".

3.6. Objective Function

The main intent of this research is to attain cost minimization while designing and sizing of the HES. The objective function that has been selected for minimization is the NPC. As an economic metric, this function takes into account a number of factors, such as capital costs (C^{NV}_{CP}), replacement costs (C^{NV}_{RP}), operating and maintenance costs (C^{NV}_{OM}), fuel consumption costs (C^{NV}_{F}), costs associated with obtaining electricity from the utility grid (C^{G}_{P}), and other pertinent factors. Furthermore, the objective function takes into consideration the revenue obtained from the sale of electricity to the power grid at a predetermined price (C^{G}_{S}) and salvage value (C^{NV}_{SL}). The mathematical expression for these financial factors is as follows:

$$NPC = C^{NV}_{CP} + C^{NV}_{RP} + C^{NV}_{OM} - C^{NV}_{SL} + C^{NV}_{F} + C^{G}_{P} - C^{G}_{S} \tag{19}$$

3.7. Constraints

The present system's optimization is limited by practical and technical constraints that must be addressed so as to provide viable solutions.

3.7.1. HES Components Constraints

The sizing of multiple components, such as N_{SP_V}, N_{Bt}, P_G, and P_M is figure out based on the electrical load demand. Consequently, constraints are there by set forth by carving out minimum and maximum boundaries for these components, which are elucidated as follows:

$$N_{SP_V} = \text{Integer},\ N^{n}_{SP_V} \le N_{SP_V} \le N^{x}_{SP_V} \tag{20}$$

$$N_{Bt} = \text{Integer},\ N^{n}_{Bt} \le N_{Bt} \le N^{x}_{Bt} \tag{21}$$

$$P_G = \text{Integer},\ P^{n}_{G} \le P_G \le P^{x}_{G} \tag{22}$$

$$P_M = \text{Integer},\ P^{n}_{M} \le P_M \le P^{x}_{M} \tag{23}$$

3.7.2. Battery Storage Capacity Boundaries

To ensure the safe operation of battery energy storage systems, it is imperative to specify both lower and upper capacity boundaries for the battery. These boundaries are articulated as follows:

$$E_{Btn} \leq E_{Bt}(t) \leq E_{Btx}$$

3.7.3. Power Reliability Constraint

The investigation in hand covers the possibility of the loss of power supply while addressing power reliability. When a user's energy demand exceeds the readily available generation, they are not able to obtain energy. Consequently, the following formula can be used to determine LPSP [20]:

$$\text{LPSP} = \sum_{t=1}^{t=8760} \frac{\text{Unserved load for one year}}{\text{Total load for one year}} \quad (24)$$

4. Description of Study Area

A set of villages in Uttar Pradesh, India's Agra district, encompasses the selected area. NASA meteorological data is used to obtain the GHI and ambient temperature data for the chosen area [52]. It can be inferred from the wind resources that there isn't enough wind in this region to generate electricity. As a result, wind is not used when modeling H_ES. This region produces 395 tons of biomass annually from agricultural waste, and 739 m^3 of biogas is produced every day. The studied region receives, on average, 5.27 kWh/m2 of solar radiation per day. The study classifies the year into three distinct seasons, each lasting four months, pursuant to the belief that temperature variations have an immense impact on energy demand. The summer season (S_S) is outlined as April through July, the moderate season (S_M) as August through November, and the winter season (S_W) as December through March. The investigating geographical area estimated daily energy consumptions for S_S, S_M, and S_W seasons are 2985.66, 2341.72, and 1266.44 kWh/day, respectively. Further, Figs. 1, 2, and 3 provides the graphic representation of the hourly electrical load demand, solar irradiation, and ambient temperature at the designated area. Moreover, the scheduling of bio-generators during peak load hours in each season is carried out in the present work. During S_S and S_M, the B_G is scheduled from 1:00 AM to 5:00 AM, while the B_M will operate from 6:00 PM to 12:00 PM in S_S and 04.00–12.00 pm in S_M, respectively. In the case of S_W, the B_G will operate in the evening from 5:00 PM to 7:00 PM, and the BM will function from 8:00 PM to 10:00 PM, respectively.

5. Techno-Economic Indexes

In this preliminary investigation, the optimization outcomes were appraised based on various techno-economic indices specified in Table 1 [41]. Additionally, the annual rates of real interest, inflation, and escalation are pegged at 0.11, 0.075, and 0.05, respectively. Beyond that, for optimizing theobjective function, a number of parameters including search agents, maximum iterations, and trial runs for TSA, HS and PSO algorithms are set as 30, 150 and 30 respectively.

6. Results and Discussion

In the present study, at first, the goal is to generate hourly forecasts for GHI and temperature for a one-year span. For this purpose, a variety of ML techniques, notably SVR, GPR, XGBT, and DT, are applied. Further, the performance of these models is compared through statistical indicators, and the most effective solution is chosen. After that, the acquired forecasted data is employed to size the HES. The sizing process of HES is accomplished by applying a recently developed technique called TSA, both with as well as without the aid of forecasted data. Added to that, a comparison has been made between the TSA, HS and PSO approaches. These algorithms undergo implementation by way of MATLAB, which runs hourly simulations for selected grid-linked HES over a year. After that, the outcomes are contrasted according to how best to minimize NPC while retaining power supply reliability.

6.1. Forecasting of Meteorological Data

Hourly forecasting for GHI and temperature has been conducted over the course of one year period using various ML techniques, including SVR, GPR, XGBT, and DT models. To expansively appraise the accuracy of the forecasted models (FMs), a scrupulous training and testing approach was adopted. First, four years of historical data for temperature and GHI from 2016 to 2019 were used to train multiple FMs. In the aftermath of training, the FMs underwent a real prediction stage, referred to as the testing phase, where they were able to forecast temperature and GHI via 2020 data. Through comparing the predicted values with the actual data and using statistical parameters namely the root mean square error (E_{RMS}), mean absolute error (E_{MA}), and correlation coefficient (R^2), the precision of each FM was checked out. Further, the statistical parameters for GHI and temperature for both the training and testing stages using ML approaches are determined and demonstrated in Table 2.

Table 2 renders it explicit that, throughout each of the training and testing phases, including temperature and GHI, the GPR model has the smallest E_{RMS}, E_{MA}, and largest R^2. As a result, HES has been developed using the

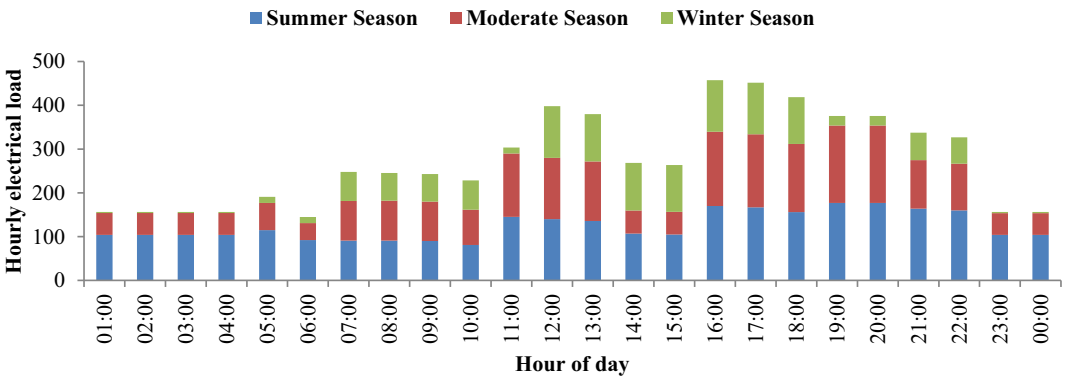


Fig. 1 Hourly electrical load demand of the designated area

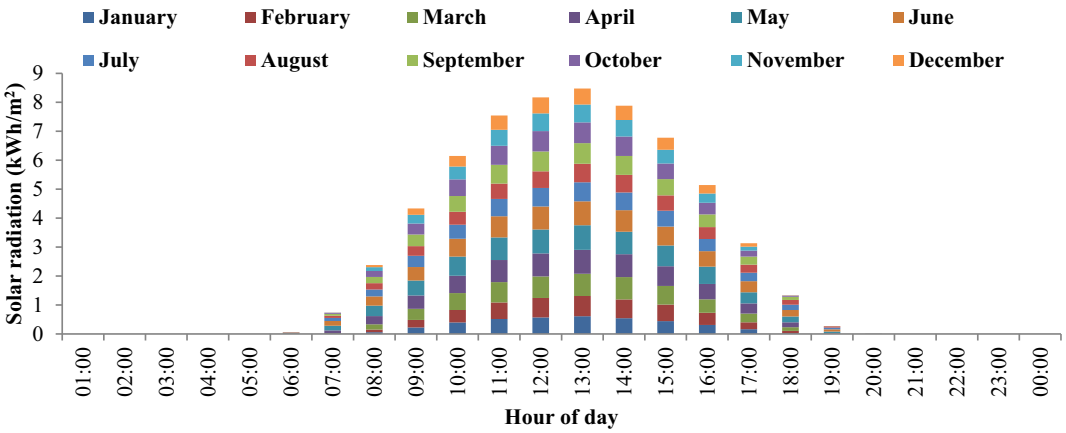


Fig. 2 Month wise solar radiation of given region

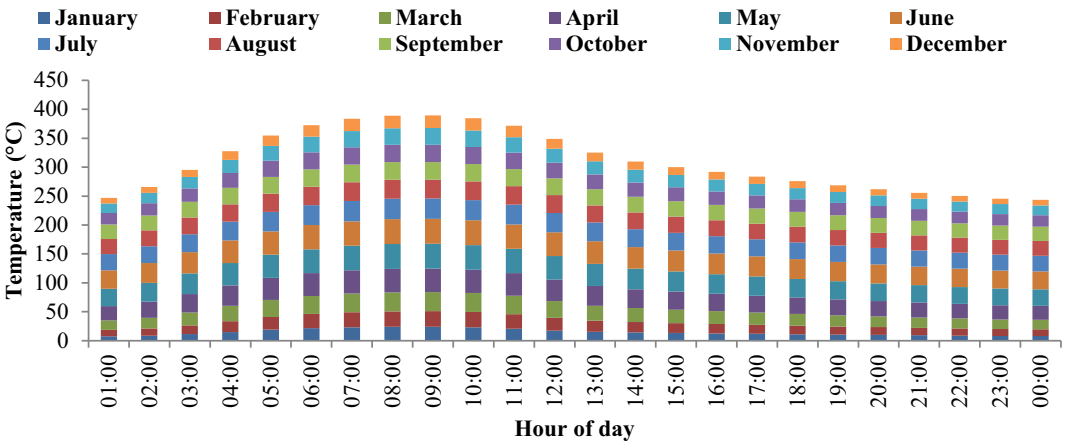


Fig. 3 Month wise temperature of given region

predicted GHI, and temperature data that were gathered through the GPR approach.

6.2. Optimization Results of Proposed HES

After acquiring the forecasted data for GHI and temperature for the chosen area, the optimal sizing of grid-linked HES, encompassing multiple R_ES, is practiced to accommodate the energy necessitates for the location at hand while utilizing the least amount of NPC and CE. TSA, a recently introduced optimization approach, was used to optimize the size of the intended HES. Likewise, to validate the findings, both the HS and PSO algorithms were applied for sizing. To facilitate this analysis, an hourly

Table 1 Economic indexes of HES components [49]

$R_E S$ system	Capacity description	Capital cost ($)	Operation and maintenance cost ($) Cost	Salvage value ($)	Fuel cost
S_{PV} system	235 W	$151.71	3.03	15.17	-
B_M system	1kW	$816.58	40.83	244.97	15.18 $/ton
B_G system	1kW	$521.72	26.08	156.52	8.09 $/ton

Table 2 Statistical parameters of ML approaches for GHI and temperature during training and testing phase

Meteorological Index	ML model	Training phase			Testing phase		
		E_{RMS}	E_{MA}	R^2	E_{RMS}	E_{MA}	R^2
GHI	SVR	0.094136684	0.051410203	0.897586	0.16992	0.1282659	0.944546
	XGBT	0.090643674	0.051363595	0.905045	0.174715	0.1299993	0.918481
	DT	0.089109341	0.051650366	0.908232	0.240235	0.1759061	0.896154
	GPR	0.098263664	0.05953362	0.888409	0.187653	0.1360788	0.945525
Temperature	SVR	0.0899271	0.008861715	0.908895	0.566462	0.0288729	0.891659
	XGBT	0.09705954	0.007940475	0.92675	0.694554	0.0577127	0.910532
	DT	0.0840803	0.009655748	0.9178016	0.261181	0.0352136	0.905154
	GPR	0.0895453	0.008216276	0.933532	0.801434	0.0305253	0.921403

simulation of selected HES was conducted in MATLAB for an entire year, with 150 iterations and 30 runs. Subsequently, a comparison of the different optimization outcomes derived through the use of these algorithms with and without forecasted data is shown in Table 3.

Upon analyzing the Table 3, it is clear that the TSA optimizer stands out as the most inexpensive option, with a minimal NPC of $519,340 and an enticing CE of 0.0606 kWh. These outcomes imply that the TSA prevails over the other selected approaches. A B_M of 85 kW, a B_G of 29 kW, a S_{PV} system of 234.7 kW (991 units), battery storage of 84 kWh (35 units), and a converter with a capacity of 100 kW are the optimal component sizes for this intended system. Additionally, some of the essential parameters incurred with the proposed grid-tied HES have also been looked at and addressed in the paragraphs that follow.

The fraction of $R_E S$ in the overall annual electricity generation facilitated by the implemented grid-tied HES is illustrated in Fig. 4. The data point out that S_{PV} panels, with an annual output of 439,410 kWh, are the predominant contributor to electricity generation. Next queued up were B_G and B_M generators that produced 45,907 kWh and 196,775 kWh annually, respectively.

Along with this, Table 4 lends a comprehensive analysis of the NPC by partitioning it into distinct categories that reflects revenue and cost components.

The data presented above reveals that the cost that stands out the most among all is the cost of hiring electricity. Under the given configuration, the total amount spent on grid purchases is $239,430, and the total amount

Table 3 Algorithms results for HES with and without forecasted data

Indices	TSA		HS		PSO	
	Without forecasted data	With forecasted data	Without forecasted data	With forecasted data	Without forecasted data	With forecasted data
N_{SP_V}	999	991	973	991	991	967
N_{Bt}	39	35	33	38	48	34
P_G(kW)	25	29	31	32	30	32
P_M(kW)	88	85	65	88	96	55
NPC ($)	520,680	519,340	527,470	524,180	533,480	532,380
CE ($/kWh)	0.0608	0.0606	0.0621	0.0613	0.0629	0.0628

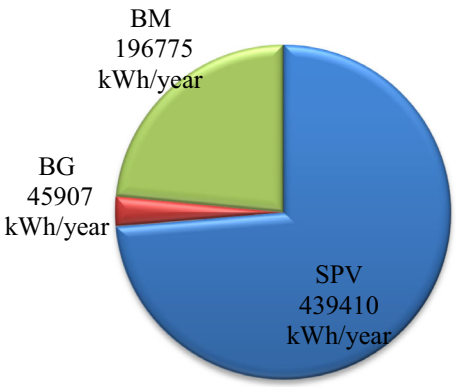


Fig. 4 Distribution of $R_E S$ in yearly electricity generation

Table 4 Cost wise distribution in NPC

S. no	Cost/revenue	Value ($)
1	C_{CP}^{NV}	137,430
2	C_{OM}^{NV}	130,520
3	C_{F}^{NV}	62,157
4	C_{SL}^{NV}	10,499
5	C_{P}^{G}	239,430
6	C_{S}^{G}	39,689

of earnings obtained from grid sales is $39,689. Further, Fig. 5 provides additional insight into the distribution of NPCs among HES components. Interestingly, B_M makes up 53% of the total, topping the shares held by S_{PV}, B_G, battery, and converter, which are 16%, 16%, 9%, and 6% of the total, respectively.

Additionally, information on the annual amount of electricity snapped up from and traded to the utility grid for each season is evaluated. It is measured that, for the S_S, 132,190 kWh were acquired from the grid, with 15,443 kWh sold back. Similarly, in the S_M, 67,567 kWh were purchased and 7240.2 kWh sold. In contrast, during the S_W, 27,358 kWh were bought and 18,603 kWh were sold to the grid. Overall, the annual energy sold and bought via the utility grid is 41,286 kWh and 229,900 kWh,

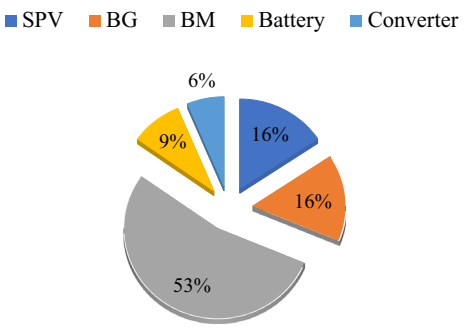


Fig. 5 NPC distribution among HES components

respectively. In response to heightened energy demand during the S_S, the HES yield a larger amount of energy in contrast to the S_M and S_W season. On the other hand, in S_W, sales and grid usage both turn down significantly, chiefly owing to lower energy demand during that time. Added to that, it is also assessed how battery input power (P_{BI}) and output power (P_{BO}) fluctuate cyclically. Specifically, P_{BO} for the S_S was the most substantial at 137,210 kWh in contrast to its P_{BI} of 22,817 kWh, indicating higher energy consumption than generation. Conversely, the S_M and S_W show higher P_{BI} values of 29,325 kWh and 39,571 kWh, respectively against their respective P_{BO} values of 80,254 kWh and 45,138 kWh, respectively, implying less energy demand. Interestingly enough, P_{BI} rises over S_W as opposed to S_S and S_M, indicating lower energy demand during this period.

7. Conclusion

This current investigation endeavors to optimize the size of HES in the far-flung territory of Agra district, India. A grid-tied system configuration was developed using forecasted data for GHI and temperature. Four ML methods—SVR, DT, GPR, and XGBT—were applied for forecasting, with GPR showcasing superior performance. As it turned out, data derived from GPR was used for sizing the HES. Further, size optimization of HES was then performed utilizing a recently developed approach, notably TSA. A comparison ensued between the TSA results and those obtained by HS and PSO. TSA encountered the optimal configuration of the HES, featuring a 234.7 kW S_{PV} system, a 29 kW B_G, an 85 kW B_M generator, and a 84 kWh battery bank with converter of 100 kW. The anticipated levelized CE and total NPC were set at $0.0606 per kWh and $519,340, respectively. The results of this research hold vitality for the development of HES in locales with reminiscent morphology. The implementation of these findings could lead to a consistent and reliable electricity supply, which would strengthen rural communities' quality of life. The result obtained from the proposed model is compared with the measured data to determine the level of uncertainty and accuracy. Since the proposed model has accuracy within the desired limit, therefore, the experimental and real-time models may be deployed at the selected site to provide electrical energy to consumers. Additionally, these research inferences line up with the aspirations of the Indian government, particularly in urging "power to all" diligence and striving to guarantee electricity supply 24 hours a day, seven days a week.

References

1. M. Thirunavukkarasu, H. Lala and Y. Sawle, Techno-economic-environmental analysis of off-grid hybrid energy systems using honey badger optimizer. Renew. Energy., *218* (2023) 119247.
2. S. Diwania, S. Agrawal and A.S. Siddiqui, Performance enhancement of single-channel glazed photovoltaic thermal module using whale optimisation algorithm and its comparative study. Int. J. Ambient Energy, *42*(4) (2021) 424–434. https://doi.org/10.1080/01430750.2018.1537937.
3. H.T. Nazloo, R. Babazadeh and T.G. Bolandi, Developing a mathematical programming model for planning and sizing of grid-connected microgrids. Renew. Energy Focus, *44* (2023) 212–222.
4. A. Gupta, S. Agrawal and Y. Pal, Performance evaluation of hybrid photovoltaic thermal thermoelectric collector using grasshopper optimization algorithm with simulated annealing. J. Solar Energy Eng., *142*(6) (2020) 1–14.
5. M. Thirunavukkarasu, H. Lala and Y. Sawle, Reliability index based optimal sizing and statistical performance analysis of stand-alone hybrid renewable energy system using metaheuristic algorithms. Alex. Eng. J., *74* (2023) 387–413.
6. G.S. Chaurasia, A.K. Singh, S. Agrawal and N.K. Sharma, A meta-heuristic firefly algorithm based smart control strategy and analysis of a grid connected hybrid photovoltaic/wind distributed generation system. Solar Energy, *150* (2017) 265–274.
7. C. Ceylan and Y. Devrim, Green hydrogen based off-grid and on-grid hybrid energy systems. Int. J. Hydrogen Energy, *48* (2023) 39084–39096.
8. D. Emad, M.A. El-Hameed and A.A. El-Fergany, Optimal techno-economic design of hybrid PV/wind system comprising battery energy storage: case study for a remote area. Energy Conver. Manag., *249* (2021) 114847.
9. H. Zahboune, S. Zouggar, G. Krajacic, P.S. Varbanov, M. Elhafyani and E. Ziani, Optimal hybrid renewable energy design in autonomous system using modified electric system cascade analysis and Homer software. Energy Conver. Manag., *126* (2016) 909–922.
10. P.P. Kumar and R.P. Saini, Optimization of an off-grid integrated hybrid renewable energy system with different battery technologies for rural electrification in India. J. Energy Storage, *32* (2020) 101912. https://doi.org/10.1016/j.est.2020.101912.
11. J. Ahmad, M. Imran, A. Khalid, W. Iqbal, S.R. Ashraf, M. Adnan, S.F. Ali and K.S. Khokhar, Techno economic analysis of a wind photovoltaic-biomass hybrid renewable energy system for rural electrification: a case study of Kallar Kahar. Energy, *148* (2018) 208–234. https://doi.org/10.1016/j.energy.2018.01.133.
12. X. Yuan, L. Heikari, J. Hirvonen, Y. Liang, M. Virtanen, R. Kosonen and Y. Pan, System modelling and optimization of a low temperature local hybrid energy system based on solar energy for a residential district. Energy Conver Manag., *267* (2022) 115918.
13. P. Anand, S.K. Bath and M. Rizwan, Renewable energy-based hybrid model for rural electrification. Int. J. Energy Technol. Policy, *15* (2019) 86–113.
14. S.M. Mahmoudi, A. Maleki and D. Rezaei Ochbelagh, Optimization of a hybrid energy system with/without considering back-up system by a new technique based on fuzzy logic controller. Energy Conver. Manag., *229* (2021) 113723.
15. A. Al-Othman, M. Tawalbeh, R. Martis, S. Dhou, M. Orhan, M. Qasim and A.G. Olabi, Artificial intelligence and numerical models in hybrid renewable energy systems with fuel cells: advances and prospects. Energy Conver. Manag., *253* (2022) 115154.
16. R. Kumar, R.A. Gupta and A.K. Bansal, Economic analysis and power management of a stand-alone wind/photovoltaic hybrid energy system using biogeography based optimization algorithm. Swarm Evol. Comput., *8* (2013) 33–43.
17. A. Maleki and F. Pourfayaz, Sizing of stand-alone photovoltaic/wind/diesel system with battery and fuel cell storage devices by harmony search algorithm. J. Energy Storage, *2* (2015) 30–42.
18. A.A. Fathy, Reliable methodology based on mine blast optimization algorithm for optimal sizing of hybrid PV-wind-FC system for remote area in Egypt. Renew. Energy, *95*(562) (2016) 367–380.
19. R. Wang and R. Zhang, Techno-economic analysis and optimization of hybrid energy systems based on hydrogen storage for sustainable energy utilization by a biological-inspired optimization algorithm. J. Energy Storage, *66* (2023) 107469.
20. P. Anand, M. Rizwan and S.K. Bath, Sizing of renewable energy based hybrid system for rural electrification using grey wolf optimization approach. IET Energy Syst. Integr., *1* (2019) 158–172.
21. B. Sharma, M. Rizwan and P. Anand, A new intelligent approach for size optimization of a renewable energy based grid connected hybrid energy system. Int. J. Numer. Model. Electron. Netw. Devices Fields, *36* (2023) 1–19. https://doi.org/10.1002/jnm.3050.
22. A.L. Bukar, C.W. Tan and K.Y. Lau, Optimal sizing of an autonomous photovoltaic/wind/ battery/diesel generator microgrid using grasshopper optimization algorithm. Sol. Energy, *188* (2019) 685–696.
23. F. Tooryan, H. HassanzadehFard, E.R. Collins, S. Jin and B. Ramezani, Optimization and energy management of distributed energy resources for a hybrid residential microgrid. J Storage Mater, *30* (2020) 101556.
24. H. Sun, A.G. Ebadi, M. Toughani, S.A. Nowdeh, A. Naderipour and A. Abdullah, Designing framework of hybrid photovoltaic-biowaste energy system with hydrogen storage considering economic and technical indices using whale optimization algorithm. Energy, *238* (2022) 121555.
25. M. Kharrich, S. Kamel, M. Abdeen, O.H. Mohammed, M. Akherraz, T. Khurshaid and S.B. Rhee, Developed approach based on equilibrium optimizer for optimal design of hybrid pv/wind/diesel/battery microgrid in Dakhla, Morocco. IEEE Access, *9* (2021) 13655–13670.
26. M. Kharrich, A. Selim, S. Kamel and J. Kim, An effective design of hybrid renewable energy system using an improved Archimedes Optimization Algorithm: a case study of Farafra, Egypt. Energy Convers. Manag., *283* (2023) 116907.
27. R. Nuvvula, E. Devaraj and K.T. Srinivasa, A comprehensive assessment of large scale battery integrated hybrid renewable energy system to improve sustainability of a smart city. Energy Sources Part A Recover. Util. Environ. Eff. (2021). https://doi.org/10.1080/15567036.2021.1905109.
28. D.X.T. George, R.E. Raj, A. Rajkumar and M.C. Mabel, Optimal sizing of solar-wind based hybrid energy system using modified dragonfly algorithm for an institution. Energy Convers. Manag., *283* (2023) 116938.
29. N. Kumar, K. Namrata and A. Samadhiya, Deterministic robust planning and probabilistic techno-economic assessment of a sector coupled community energy system. Adv. Theory Simul., *5* (2022) 2100639. https://doi.org/10.1002/adts.202100639.
30. A.A.Z. Diab, H.M. Sultan, I.S. Mohamed, O.N. Kuznetsov and T.D. Do, Application of different optimization algorithms for optimal sizing of PV/wind/diesel/battery storage stand-alone hybrid microgrid. IEEE Access, *7* (2019) 119223–119245.
31. A. Tabak, E. Kayabasi, M.T. Guneser and M. Ozkaymak, Grey wolf optimization for optimum sizing and controlling of a PV/WT/BM hybrid energy system considering TNPC, LPSP, and LCOE concepts. Energy Sour. Part Recov Environ Eff, *44*(1)

(2022) 1508–1528. https://doi.org/10.1080/15567036.2019.1668880.
32. S. Divya, M.K. Paramathma, A. Sheela and S.D. Kumar, Hybrid renewable energy source optimization using black widow optimization techniques with uncertainty constraint. Meas. Sens., *31* (2024) 100968.
33. M. Gupta and A. Bhargava, Optimal design of hybrid renewable-energy microgrid system: a techno–economic–environment–social– reliability perspective. Clean Energy, *8*(1) (2024) 66–83. https://doi.org/10.1093/ce/zkad069.
34. T. Ramesh and K. Balachander, Optimizing hybrid renewable energy systems with integrated electric vehicle using a hybrid approach. J. Energy Storage, *89* (2024) 111655.
35. E.I.M.B. Messini, Y. Bourek, C. Ammari and A. Pesyridis, The integration of solar-hydrogen hybrid renewable energy systems in oil and gas industries for energy efficiency: optimal sizing using Fick's Law optimisation. Algorithm Energy Conver. Manag., *308* (2024) 118372.
36. W.K. Yemeli, F.C.V. Fohagui, C.A.W. Ngouleu and G. Tchuen, An effective sizing and sensitivity analysis of a hybrid renewable energy system for household, multi-media and rural healthcare centres power supply: a case study of Kaele Cameroon. Int. J. Hydrog. Energy, *49* (2024) 1321–1359.
37. M. Lei, L. Shiyan, J. Chuanwen, L. Hongling and Z. Yan, A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev, *13*(4) (2009) 915–920. https://doi.org/10.1016/j.rser.2008.02.002.
38. J.P.S. Catalao, H.M.I. Pousinho and V.M.F. Mendes, Hybrid wavelet PSO-ANFIS approach for short-term wind power forecasting in Portugal. IEEE Trans. Sustain. Energy (2010). https://doi.org/10.1109/TSTE.2010.2076359.
39. F.O. Hocaoglu, O.N. Gerek and M. Kurban, The effect of model generated solar radiation data usage in hybrid (wind–PV) sizing studies. Energy Convers. Manag., *50*(12) (2009) 2956–2963. https://doi.org/10.1016/j.enconman.2009.07.011.
40. A. Tascikaraoglu, M. Uzunoglu and B. Vural, The assessment of the contribution of short-term wind power predictions to the efficiency of stand-alone hybrid systems. Appl. Energy, *94* (2012) 156–165. https://doi.org/10.1016/j.apenergy.2012.01.017.
41. M. Khalid and A.V. Savkin, A model predictive control approach to the problem of wind power smoothing with controlled battery storage. Renew. Energy, *35*(7) (2010) 1520–1526. https://doi.org/10.1016/j.renene.2009.11.030.
42. W. Zhang, A. Maleki, M.A. Rosen and J. Liu, Sizing a stand-alone solar-wind-hydrogen energy system using weather forecasting and a hybrid search optimization algorithm. Energy Convers Manag., *180* (2019) 609–621. https://doi.org/10.1016/j.enconman.2018.08.102.
43. A.A. Medina-Santana and L.E. Cárdenas-Barrón, Optimal design of hybrid renewable energy systems considering weather forecasting using recurrent neural networks. Energies, *15*(23) (2022) 9045. https://doi.org/10.3390/en15239045.
44. A. Maleki, M.G. Khajeh and M.A. Rosen, Weather forecasting for optimization of a hybrid solar-wind–powered reverse osmosis water desalination system using a novel optimizer approach. Energy, *114* (2016) 1120–1134. https://doi.org/10.1016/j.energy.2016.06.134.
45. R.A. Gupta, R. Kumar and A.K. Bansal, BBO-based small autonomous hybrid power system optimization incorporating wind speed and solar radiation forecasting. Renew. Sustain. Energy Rev., *41* (2015) 1366–1375. https://doi.org/10.1016/j.rser.2014.09.017.
46. S. Kaur, L.K. Awasthi, A.L. Sangal and G. Dhiman, Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization. Eng. Appl. Artif. Intell., *90* (2020) 103541.
47. I. Alsaidan, M. Bilal, M. Alaraj, M. Rizwan and F.M. Almasoudi, A novel EA-based techno–economic analysis of charging system for electric vehicles: a case study of Qassim Region, Saudi Arabia. Mathematics, *2023* (2023) 1–31. https://doi.org/10.3390/math11092052.
48. P. Anand, V.K. Kamboj, M. Alaraj, M. Rizwan and I.J. Mwakitalima, Optimal sizing of hybrid energy system using random exploratory search centred harris hawks optimizer with improved exploitation capability. Math. Prob. Eng., *2022* (2022) 1–22. https://doi.org/10.1155/2022/5348017.
49. P. Anand, M. Rizwan and S.K. Bath, Size optimization of RES based grid connected hybrid power system using harmony search algorithm. Int. J. Energy Technol. Policy, *16* (2020) 238–276.
50. B. Sharma, M. Rizwan, P. Anand, Optimal sizing of a grid connected biomass/biogas/SPV system for rural electrification, Proceedings of 3rd international conference on machine learning advances in computing renewable energy and communication: lecture notes in electrical engineering, **915** (2022) 507–514. https://doi.org/10.1007/978-981-19-2828-4_46.
51. B. Sharma, M. Rizwan and P. Anand, Feasibility and techno-economic analysis of renewable energy based hybrid system for rural community of India. J. Phys. Conf. Ser., *2570*(1) (2023) 1–11.
52. https://www.nrel.gov/gis/solar.html. Accessed November 20, 2023.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Strategic Pseudo-Goal Perturbation for Deadlock-Free Multi-Agent Navigation in Social Mini-Games

Abhishek Jha*
Department of Mechanical Engineering
Delhi Technological University
New Delhi, India
aplusj247@gmail.com
*Corresponding author

Tanishq Gupta
Department of Mechanical Engineering
Delhi Technological University
New Delhi, India
gtanishq9@gmail.com

Sumit Singh Rawat
Department of Mechanical Engineering
Delhi Technological University
New Delhi, India
singhrawat.sumit@gmail.com

Girish Kumar
Department of Mechanical Engineering
Delhi Technological University
New Delhi, India
girishkumar@dce.ac.in

2024 9th International Conference on Control and Robotics Engineering (ICCRE) | 979-8-3503-7269-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCRE61448.2024.10589836

***Abstract*—This work introduces a Strategic Pseudo-Goal Perturbation (SPGP) technique, a novel approach to resolve deadlock situations in multi-agent navigation scenarios. Leveraging the robust framework of Safety Barrier Certificates, our method integrates a strategic perturbation mechanism that guides agents through social mini-games where deadlock and collision occur frequently. The method adopts a strategic calculation process where agents, upon encountering a deadlock select a pseudo goal within a predefined radius around the current position to resolve the deadlock among agents. The calculation is based on controlled strategic algorithm, ensuring that deviation towards pseudo-goal is both purposeful and effective in resolution of deadlock. Once the agent reaches the pseudo goal, it resumes the path towards the original goal, thereby enhancing navigational efficiency and safety. Experimental results demonstrates SPGP's efficacy in reducing deadlock instances and improving overall system throughput in variety of multi-agent navigation scenarios.**

***Index Terms*—Strategic Perturbation, Deadlock Avoidance, Multi-Agent Navigation, Social Mini-Games**

I. Introduction

In the domain of control and robotics, multi-agent navigation poses significant challenges, particularly when multiple autonomous agents operate within the same environment. These challenges are compounded in scenarios where agents must navigate social spaces and collisions among agents deemed to happen which are referred to as "social mini-games"—such as doorways, intersections, and L-corners. These environments necessitate not only collision avoidance but also efficient coordination among agents to prevent deadlock situations, where agents halt progress due to mutual obstruction [1]–[4].

Collision avoidance in multi-agent systems is a well-studied area, with numerous algorithms developed to enable autonomous agents to navigate without incidents [5]–[7]. But, the challenges of social mini-games introduces complex interaction dynamics which is not fully solved by the existing collision avoidance techniques. In these scenarios, agents often encounter deadlock due to the limited space for the agents to navigate and the symmetry found in the environment [8], [9]. Such deadlocks significantly restricts the flow of movement, leading to low efficiencies and performance.

Our method, Strategic Pseudo-Goal Perturbation (SPGP), addresses the challenge of deadlock in multi-agent navigation within social mini-games. SPGP innovates by combining Safety Barrier Certificates with a mechanism that, upon detecting a potential deadlock, strategically perturbs agents towards pseudo goals. These pseudo goals are generated based on a strategic algorithm that enforces efficient deviation from the agent's original path, facilitating the resolution of the deadlock. Once the agent reaches its pseudo goal, it recalculates its path towards the original goal, thus minimizing errors and increasing the efficiencies in this process. By concentrating on strategic perturbation rather than other techniques, SPGP offers solution to resolve deadlock in complex social navigation scenarios. This perturbation technique not only the efficiency of multi-agent navigation but also contributes to the improvement of autonomous agent coordination in constrained environments. Fig. 1 depicts the overview of SPGP method where navigation starts then deadlock ios detected and finally deadlock is resolved using the strategic perturbation.

This study lays contributions to the field of multi-agent navigation and deadlock avoidance in social mini-game scenarios. The key contributions are as follows:

- Introduced a Strategic Pseudo-Goal Perturbation (SPGP) method, a novel approach to generate deadlock-free situa-

979-8-3503-7269-4/24/$31.00 ©2024 IEEE

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:26:38 UTC from IEEE Xplore. Restrictions apply.

tions in multi-agent navigation. This method strategically perturbs agents towards pseudo goals.
- Through iterative analysis, explored the variation of the perturbation radius and its effect on the effectiveness of deadlock resolution, providing insights into optimal perturbation strategies for varying environmental complexities and agent densities.
- Thoroughly tested the SPGP method across various complex social mini-games scenarios. These experiments depicts the flexibility and adaptability of our method, which shows its effectiveness in a wide range of common social navigation challenges.

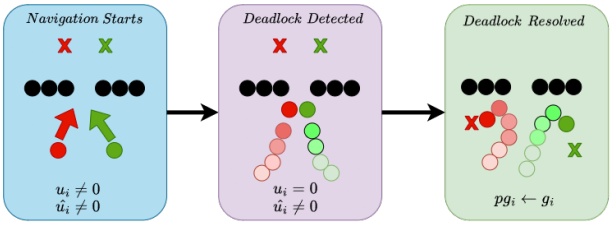


Fig. 1: Overview of Strategic Pseudo-Goal Perturbation

II. Related Work

A. Safety Based Methods

The ORCA framework [10], along with its non-holonomic extension [11], shows provable safety in multi-agent navigation through single-integrator systems. This method efficiently generates optimal collision-free velocities by using linear programming to navigate constraints represented as half-planes in velocity space. While originally designed for holonomic systems, the framework's also been extended for non-holonomic constraints, aiming to impose minimal deviation from preferred velocities. However, the construction of half-planes in ORCA can lead to deadlocks, marking its primary limitation [12]. Ensuring safety in systems with double-integrator dynamics is more challenging, with safety largely influenced by the system's planning frequency. For instance, the NH-TTC framework [13] employs gradient descent to optimize a cost function that combines position with a time-to-collision component, which helps to prevent imminent collisions. NH-TTC's safety assurances improve as planning frequency increases towards infinity. Another set of methods uses the shortest path in the multi-agent network to reach the goal position in a complex environment [29], [30]. Similarly, the model predictive control (MPC) method [8] underscores safety as a function of both planning frequency and the extent of the time horizon. Further, Control Barrier Functions (CBFs) [15], [16] offer a method to design controllers ensuring safety by maintaining the forward invariance property in the set, ensuring that if an agent starts within a safety set, it remains within this set indefinitely, effectively preventing it to move away from the safe zone.

B. Learning Based Methods

The integration of conventional navigation techniques with machine learning represents a fast-evolving research space on learning-based motion planning. Among the various techniques, deep reinforcement learning (DRL) and trajectory prediction stand out for their success in multi-agent planning and navigation. DRL is applied to generate navigation policies through simulation, including for scenarios involving multiple robots in social mini-games. For instance, [17] describes a decentralized collision avoidance system for multi-robot scenarios using DRL and local sensory inputs. Another example is CADRL [18], which optimizes social robot navigation by employing a sparse reward system, with extensions like LSTM-based action selection in response to the proximity of other robots [19]. On the other hand, algorithms incorporating trajectory prediction focus on forecasting future robot positions to navigate dynamically among moving obstacles effectively.

Another avenue is Imitation Learning (IL) which is a machine learning strategy where an agent learns behaviors by mimicking expert demonstrations, typically faster than conventional optimization methods. These demonstrations often include state input pairs from expert policy executions. The main IL methodologies are Behavior Cloning (BC) [20], [21], which uses supervised learning to directly acquire an imitative policy; Inverse Reinforcement Learning (IRL) [22], which deduces a reward function from demonstrations to facilitate policy learning via Reinforcement Learning (RL) [23]; and approaches utilizing generative models [24]. Despite their efficiencies, IL approaches [25], [26] commonly face limitations in encoding state/safety and input constraints, presuming access to expert action data during demonstrations and struggling to adapt to scenarios beyond the expert demonstrations' distribution.

C. Deadlock Resolution Methods

Recent studies have explored structured perturbation strategies, such as the right-hand rule [27] and clockwise rotation [15], to enhance performance beyond random perturbation methods, even providing formal optimality guarantees. However, these approaches are constrained by a preset ordering, which restricts their applicability, often to scenarios with no more than three agents. Another different category of deadlock resolution techniques adopts priority and scheduling mechanisms which are generally found in managing intersection in autonomous agents research [28]. Notable among these protocols are First Come First Served (FCFS), reservation, and auction systems. Specifically, this method [29] prioritizes agents by their sequence of arrival at intersections. While straightforward to apply, this method may result in extended waiting periods and increased congestion when numerous vehicles converge on an intersection at the same time.

III. Preliminaries

In this section, we lay the basis for our contributions by discussing the foundational concepts of Control Barrier Functions (CBF), Safety Barrier Certificates (SBC) and Deadlock

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:26:38 UTC from IEEE Xplore. Restrictions apply.

conditions which form the basis for ensuring safety in multi-agent navigation.

A. Control Barrier Functions

Control Barrier Functions (CBFs) provides assurance of the forward invariance property of the safe set for the agents [15], [17].

$$\dot{x} = f(x) + g(x)u, \tag{1}$$

where $x \in \mathbb{R}^n$ represents the state and $u \in \mathbb{R}^m$ represents the control input, a set $\mathcal{C} \subseteq \mathbb{R}^n$ is said to be forward invariant if, for the initial condition $x(0) \in \mathcal{C}$, we have $x(t) \in \mathcal{C}$ for all $t \geq 0$. The set $\mathcal{C}$ is defined as the zero-superlevel set of a continuously differentiable function $h : \mathbb{R}^n \to \mathbb{R}$, such that

$$\mathcal{C} = \{x \in \mathbb{R}^n \mid h(x) \geq 0\}. \tag{2}$$

The time derivative of $h(x)$ around the trajectories of the agent's system is given by

$$\frac{dh(x)}{dt} = \frac{\partial h}{\partial x} \cdot \frac{dx}{dt} = \frac{\partial h}{\partial x} \cdot (f(x) + g(x)u), \tag{3}$$

or, using the Lie derivative form,

$$\frac{dh(x)}{dt} = L_f h(x) + L_g h(x)u. \tag{4}$$

A function $h(x)$ is a CBF for the set $\mathcal{C}$ if there exists an extended class-$\mathcal{K}$ function κ such that for all $x \in \mathbb{R}^n$,

$$\sup_{u \in \mathbb{R}^m} [L_f h(x) + L_g h(x)u + \kappa(h(x))] \geq 0, \tag{5}$$

where $L_f h(x)$ and $L_g h(x)$ denote the Lie derivatives of h along f and g, respectively. For $h(x)$ to serve as a CBF, it must satisfy the inequality

$$L_f h(x) + L_g h(x)u + \gamma h^3(x) \geq 0, \quad \forall x \in \mathbb{R}^n, \tag{6}$$

for some extended class-$\mathcal{K}$ function $\gamma h^3(x)$, assuring the system resides within the safe set $\mathcal{C}$ for every future timesteps. where γ represents the controlling parameter and $h(x)$ represents the barrier function.

B. Safety Barrier Certificates

Safety Barrier Certificates (SBCs) are designed for enforcing collision avoidance among various agents in a multi-agent system [15]. For agents i and j, with dynamics modeled as double integrator, the relative velocity and position vectors are denoted as Δp_{ij} and Δv_{ij} respectively. To ensure a safe distance D_s is maintained, the joint constraint for safety barrier of agents $i \neq j$ is expressed in equation 7:

$$A_{ij}u \leq b_{ij}, \tag{7}$$

where A_{ij} and b_{ij} are defined as follows:

- $A_{ij} = \begin{bmatrix} 0 \dots -\Delta p_{ij}^\top \dots \Delta p_{ij}^\top \dots 0 \end{bmatrix}$ with the corresponding states of agents i and j inserted in the matrix,
- $b_{ij} = \gamma h_{ij}^3 + \frac{\Delta v_{ij}^\top \Delta p_{ij}}{2(\alpha_i + \alpha_j)} - \frac{(\Delta v_{ij}^\top \Delta p_{ij})^2}{2\|\Delta p_{ij}\|^2} + \frac{\|\Delta v_{ij}\|^2}{2}$,
- $\gamma > 0$ is a controller parameter that regulates the approach to the safety boundary.

These constraints ensure that agent i's control input u_i and agent j's control input u_j avoid collisions by satisfying the safety barrier certificates, integrated within a quadratic programming (QP) framework for real-time applications.

Note: In the definition of b_{ij}, γ, h_{ij}, Δp_{ij}, and Δv_{ij} correspond to the controller parameter, the CBF function for agents i and j, relative position vector, and the relative velocity vectors among the two agents, respectively. The symbols α_i and α_j represent the acceleration constraints of agents i and j.

C. Deadlock Condition

Deadlock scenario in multi-agent navigation systems primarily arise due to the formation of symmetries in the environment that may lead to conflicts between agents. Such symmetries often manifest in scenarios where multiple agents, following identical or mirror-like navigation policies, encounter each other in narrow passages, intersections, or areas with restricted maneuverability. The mathematical expression for identifying a deadlock situation hinges on the actual control inputs u_i and the nominal control inputs $\hat{u}_i$. The nominal control inputs represent the intended direction and magnitude of movement in the absence of other agents in the given scenario. In contrast, the actual control inputs are adjusted based on the real-time navigation strategy, incorporating obstacle avoidance and agent-to-agent spacing requirements. Fig. 2 depicts the actual and nominal scenario along with the behaviour of controls which enforces deadlock to happen, equation 8 shows the condition for the deadlock to happen.

$$\text{Deadlock Condition:} \quad u_i = 0 \quad \text{and} \quad \hat{u}_i \neq 0 \tag{8}$$

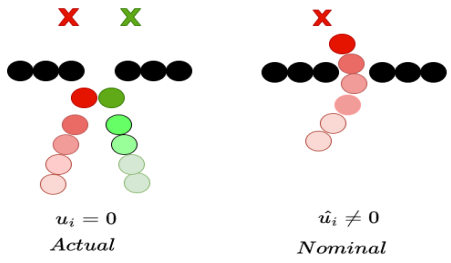


Fig. 2: Condition for deadlock among agents to occur

IV. Proposed Methodology

We introduce the Strategic Pseudo-Goal Perturbation (SPGP) framework, which is specifically designed to resolve deadlock in a centralized multi-agent environment through the use of pseudo goals. SPGP synthesizes strategic navigation with robust safety mechanisms to negotiate complex social interactions between agents.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:26:38 UTC from IEEE Xplore. Restrictions apply.

A. Mathematical Formulation

In the Strategic Pseudo-Goal Perturbation (SPGP) framework, we analyse the problem with efficiently navigating a given set of autonomous agents $\mathcal{A} = \{1, 2, \ldots, N\}$ towards their respective targets within a given environment. Each agent i, starting from an initial position $\mathbf{p}_i(t)$, is tasked with reaching a designated goal $\mathbf{g}_i$. The SPGP framework aims to optimize the planned path so as to ensuring deadlock and collision avoidance and generating optimal path for the agents. This is achieved by generating the agent's trajectories using Safety Barrier Certificates (SBCs), which enforces a safety certificates using CBFs around each agent to prevent inter-agent collisions.

1) Safety Barrier Certificates: The SBCs are important in the SPGP's collision avoidance strategy, imposing constraints that dynamically adapt to the changing states and control of the agents. Given any two agents i and j, the safety constraints are formulated in the equation 9:

$$h_{ij}(t) = \|\mathbf{p}_i(t) - \mathbf{p}_j(t)\| - (r_i + r_j) \geq 0, \tag{9}$$

where $h_{ij}(t)$ denotes the euclidean distance between the two agents, and r_i and r_j represents their respective safety radius. This function ensures that the agents maintain a separation distance which should be greater than the sum of their respective safety radius, effectively establishing a collision-free path for the agents.

B. Pseudo-Goal Selection Strategy

When the agents actual controls are below a certain threshold ($\mathbf{u}_t$) then the deadlock occurs in the simulation. When deadlock occurs the pseudo-goal selection strategy plays an instrumental role in alleviating deadlock conditions that materialize when paths of agents intersect, leading to potential immobilization. This strategy ingeniously circumvents such predicaments by allocating an interim waypoint, designated as a pseudo-goal ($\mathbf{pg}_i$), towards which an agent can momentarily divert its course.

The framework develops the navigable region $\mathcal{B}(\mathbf{p}_i, \delta)$ as a circular section, the centre is positioned on the agent's existing position, with radius δ defining its extent. Within this area, a pseudo-goal is selected based on generating a random point x, y that aims to maximize the distance from surrounding agents, in alignment with the safety objectives for barrier certificates . The given procedure can be mathematically represented by the equation 10:

$$\mathbf{pg}_i = \mathbf{p}_i + \delta \times \begin{bmatrix} \cos(2\pi \times \mathrm{x}) \\ \sin(2\pi \times \mathrm{y}) \end{bmatrix}, \tag{10}$$

Fig. 3 depicts the strategic pseudo goal selection in a circle to resolve deadlock among the agents and enforce the maximum separation between agents.

Upon the identification of a feasible pseudo-goal, the agent updates its control inputs to navigate towards this new temporary target.The resulting trajectory is then a smooth, deadlock-free path that respects the dynamic constraints of the agent and is robust to uncertainties in the environment. As agents move towards their pseudo-goals, the path adjustment mechanism continuously updates the planned path to resolve the changing states of other agents and potential new obstacles in this process.

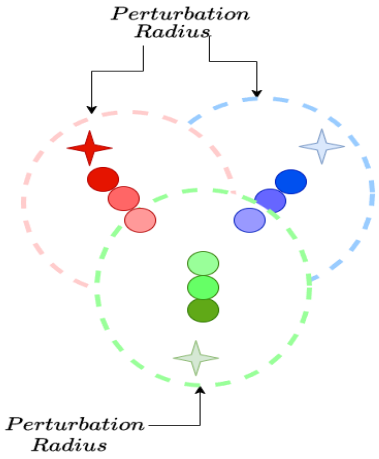


Fig. 3: Pseudo goal selection to resolve deadlock and enforce separation between agents

After reaching the pseudo-goal, agents resets their navigation systems back towards the primary target location. Algorithm 1 summarizes the Strategic Pseudo-Goal Perturbation (SPGP) algorithm and it's working to resolve the possible deadlocks.

C. Optimization Problem Formulation

The mathematical formulation of this optimization problem uses the core of Quadratic Programming (QP) expressed in equation 11. This approach tries to minimize the cost function $C(\mathbf{u})$, which serves as the main objective of our optimization. The objective function, represented in the equation 11, aims to reduce the squared deviation for the actual control inputs, $\mathbf{u}_i$, out of the nominal scenario, $\hat{\mathbf{u}}_i$. This minimization of cost function leads in achieving the scenario to mirror the nominal controller's trajectory, ensuring that each state remains as close to the intended path as feasibly possible, thus maximizing the accuracy and efficiency.

$$\begin{aligned} \mathbf{u}^* = \underset{\mathbf{u}\in\mathbb{R}^{2N}}{\operatorname{argmin}}\, C(\mathbf{u}) = \sum_{i=1}^{N} \|\mathbf{u}_i - \hat{\mathbf{u}}_i\|^2 \\ \text{s.t.} \\ A_{ij}\mathbf{u} \leq b_{ij}, \quad \forall i \neq j, \\ \|\mathbf{u}_i\|_2 \leq \alpha_i, \quad \forall i \in \mathcal{A}. \end{aligned} \tag{11}$$

This formulation considers the necessary constraints which includes $A_{ij}\mathbf{u} \leq b_{ij}$ condition that represents a safety barrier, which makes sure that each agent maintains a safe distance from its neighbouring agents, thereby reducing potential collisions. Furthermore, the condition $\|\mathbf{u}_i\|_2 \leq \alpha_i$ imposes a maximum limit on the magnitude of control inputs, confining all the agents within the practical and feasible values of dynamic capabilities .

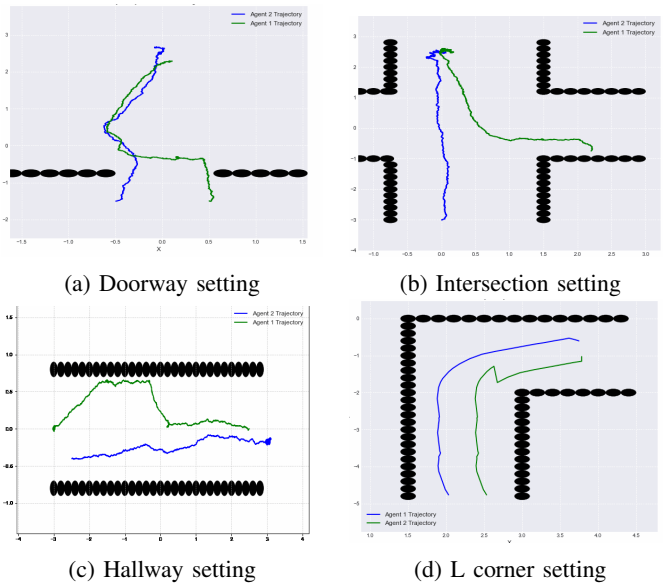
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:26:38 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Strategic Pseudo-Goal Perturbation (SPGP)

Initialize $\mathbf{p}_i(t)$, $\mathbf{g}_i$, $\mathbf{u}_t$
$u_i \leftarrow$ actual controls
$\hat{u_i} \leftarrow$ nominal controls
Deadlock_Flag $\leftarrow$ False
temp_var $\leftarrow \mathbf{g}_i$
Run SBC
if $u_i \leq u_t$ **and** $\hat{u_i} > u_t$ **then**
    Deadlock_Flag $\leftarrow$ True
end if
if Deadlock_Flag = True **then**
    $\mathbf{g}_i \leftarrow \mathbf{pg}_i$
end if
Update $\mathbf{p}_i(t)$
$x, y \leftarrow \mathbf{pg}_i$
$\mathbf{g}_i \leftarrow$ temp_var
Run SBC Again

V. Results and Discussions

This section shows outcomes and comparative study of our Strategic Pseudo-Goal Perturbation (SPGP) framework against a suite of established methodologies within the multi-agent navigation domain. Central to our evaluation is the examination of Success rate (SR), average velocity change (ΔV) and path deviation across four challenging social mini-games scenarios. The optimization of agent trajectories, facilitated by Quadratic Programming (QP), seeks to minimize deviations from nominal controls, effectively mirroring the nominal controller's behavior.



(a) Doorway setting (b) Intersection setting

(c) Hallway setting (d) L corner setting

Fig. 4: Planned trajectories of two agent in social navigation settings using strategic perturbation

Table I shows the results for the SPGP and other baseline methods for Deadlock Resolution, Average change in velocity and Path difference. Due to random nature of our algorithm we carried out each scenario experiment 10 times and recorded the mean values and standard deviation values of the scores. The perturbation radius is taken as 1m and safety radius is taken as 0.2m for all agents. Our approach is meticulously benchmarked against counterparts such as SBC, ORCA and NH-TTC, Success rate measures the efficacy of reaching the goal position without and deadlock and collision, average change in velocity for each scenario is shown which should be minimal while navigation and path deviation is the difference in trajectory of actual and nominal case for the concerned scenario. It is measured by calculating the Hausdorff distance between actual and nominal trajectory. Notably, the SPGP framework depicts higher performance in achieving success as compared with different other methods, further SPGP also yields positive results in minimizing velocity changes and path deviations, thereby asserting its efficacy in ensuring smooth and efficient agent navigation. The trajectory for two agent scenario is depicted in Fig. 4. For doorway and intersection scenarios, the goal position for both agents were same. The SPGP successfully reaches the specified goal position without any collision and deadlock by enforcing the perturbation strategy. By ablation study, we concluded that SPGP is easily scalable upto 8 agents in the doorway setting, 10 agents in the intersection and hallway setting, and 5 agents in the L-corner setting however it was noted that on increasing the number of agents the time to goal for each agent increases which increases the simulation time and the maximum number of agent that each setting can handle vastly depends on the geometric properties of the social min-game scenarios.

TABLE I: Performance Comparison Across Different Navigation Strategies

Scenario	Method	SR(%)	Avg. ΔV	Path Deviation
Doorway	SBC [15]	0.00 ± 0.00	0.10 ± 0.01	0.16 ± 0.02
	ORCA [10]	50.00 ± 0.00	0.100 ± 0.00	0.000 ± 0.00
	NH-TTC [13]	0.00 ± 0.00	0.2 ± 0.00	0.160 ± 0.00
	SPGP (Ours)	$\mathbf{100 \pm 0.00}$	$\mathbf{0.03 \pm 0.02}$	$\mathbf{0.187 \pm 0.02}$
Intersection	SBC [15]	0.00 ± 0.00	0.19 ± 0.01	0.08 ± 0.005
	ORCA [10]	50.00 ± 0.00	0.250 ± 0.00	0.89 ± 0.00
	NH-TTC [13]	50.00 ± 0.00	0.02 ± 0.00	0.47 ± 0.00
	SPGP (Ours)	$\mathbf{100 \pm 0.00}$	$\mathbf{0.034 \pm 0.01}$	$\mathbf{0.4 \pm 0.14}$
Hallway	SBC [15]	0.00 ± 0.00	0.005 ± 0.003	0.47 ± 0.05
	ORCA [10]	100.00 ± 0.00	0.110 ± 0.00	1.990 ± 0.00
	NH-TTC [13]	100.00 ± 0.00	0.02 ± 0.00	0.35 ± 0.00
	SPGP (Ours)	$\mathbf{100 \pm 0.00}$	$\mathbf{0.05 \pm 0.01}$	$\mathbf{0.32 \pm 0.32}$
L corner	SBC [15]	25.00 ± 25.00	0.11 ± 0.01	0.65 ± 0.02
	ORCA [10]	50.00 ± 0.00	0.28 ± 0.00	0.99 ± 0.00
	NH-TTC [13]	100.00 ± 0.00	0.02 ± 0.00	0.32 ± 0.00
	SPGP (Ours)	$\mathbf{100 \pm 0.00}$	$\mathbf{0.026 \pm 0.02}$	$\mathbf{0.69 \pm 0.35}$

Another experiment is performed to analyse the the effect on performance with variation of perturbation radius. In SPGP based on the perturbation radius the agent's perturbs from it's designated path to avoid deadlock and this effects the overall time in which the given agent's arrives at the goal position. To analyse this we measured the variation of makespan with different perturbation radius. The makespan value shows the time or timesteps required for the given agent to arrive at the goal position.

Fig 5 shows the change in makespan as the perturbation radius is varied. It is observed that on decreasing the perturbation radius the makepsan decreases, this happens because the agent's pseudo goal selection is confined to a circle of small radius. Also it was observed that after perturbation radius of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:26:38 UTC from IEEE Xplore. Restrictions apply.

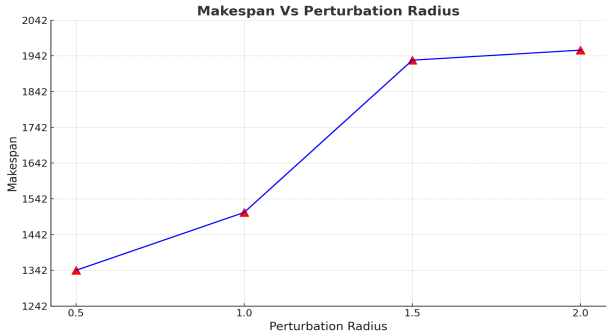


Fig. 5: Effect of different perturbation radius on the makespan of agent.

1.5m their is not much increase in makespan indicating that agent is not selecting too far away pseudo goal position so as to enforce distance between the deadlocked agents.

VI. Conclusion and Future Work

The study introduced a unique Strategic Pseudo-Goal Perturbation (SPGP) method, integrating Safety Barrier Certificates (SBCs) and a strategic pseudo-goal strategy to improve multi-agent navigation efficiency in complex social scenarios. Our findings, validated through diverse simulation scenarios, underscore SPGP's effectiveness in reducing makespan and enhancing path optimization. Future endeavors will focus on incorporating machine learning for predictive congestion management, extending the framework to heterogeneous agent systems for advanced coordination, and empirical validation in real-world settings to improve its application. Another extension of this work can be to integrate the SPGP framework with multi-agent reinforcement learning approaches and evaluate it on different social mini-game scenarios to check the method's efficacy for deadlock avoidance.

References

[1] Rasheed, Ammar Abdul Ameer, Mohammed Najm Abdullah, and Ahmed Sabah Al-Araji. "A review of multi-agent mobile robot systems applications." International Journal of Electrical and Computer Engineering 12, no. 4 (2022): 3517-3529.

[2] Zhou, Chengmin, Bingding Huang, and Pasi Fränti. "A review of motion planning algorithms for intelligent robots." Journal of Intelligent Manufacturing 33, no. 2 (2022): 387-424.

[3] Antonyshyn, Luke, Jefferson Silveira, Sidney Givigi, and Joshua Marshall. "Multiple mobile robot task and motion planning: A survey." ACM Computing Surveys 55, no. 10 (2023): 1-35..

[4] Liu, Yuanchang, and Richard Bucknall. "A survey of formation control and motion planning of multiple unmanned vehicles." Robotica 36, no. 7 (2018): 1019-1047.

[5] Huang, Sunan, Rodney Swee Huat Teo, and Kok Kiong Tan. "Collision avoidance of multi unmanned aerial vehicles: A review." Annual Reviews in Control 48 (2019): 147-164.

[6] Raibail, Mehak, Abdul Hadi Abd Rahman, Ghassan Jasim AL-Anizy, Mohammad Faidzul Nasrudin, Mohd Shahrul Mohd Nadzir, Nor Mohd Razif Noraini, and Tan Siok Yee. "Decentralized multi-robot collision avoidance: A systematic review from 2015 to 2021." Symmetry 14, no. 3 (2022): 610.

[7] Singh, Priyanka, Ritu Tiwari, and Mahua Bhattacharya. "Navigation in Multi Robot system using cooperative learning: A survey." In 2016 International Conference on Computational Techniques in Information and Communication Technologies (ICCTICT), pp. 145-150. IEEE, 2016.

[8] Y. Chen, M. Guo, and Z. Li, "Recursive feasibility and deadlock resolution in mpc-based multi-robot trajectory generation," arXiv preprint arXiv:2202.06071, 2022.

[9] J. Grover, C. Liu, and K. Sycara, "Why does symmetry cause deadlocks?," IFACPapersOnLine, vol. 53, no. 2, pp. 9746–9753, 2020.

[10] Van Den Berg, Jur, Jamie Snape, Stephen J. Guy, and Dinesh Manocha. "Reciprocal collision avoidance with acceleration-velocity obstacles." In 2011 IEEE International Conference on Robotics and Automation, pp. 3475-3482. IEEE, 2011.

[11] J. Alonso-Mora, A. Breitenmoser, M. Rufli, P. Beardsley, and R. Siegwart, "Optimal reciprocal collision avoidance for multiple non-holonomic robots," in Distributed Autonomous Robotic Systems: The 10th International Symposium, pp. 203–216, Springer, 2013

[12] S. Dergachev and K. Yakovlev, "Distributed multi-agent navigation based on reciprocal collision avoidance and locally confined multi-agent path finding," in 2021 IEEE 17th International Conference on Automation Science and Engineering (CASE), pp. 1489–1494, IEEE, 2021.

[13] B. Davis, I. Karamouzas, and S. J. Guy, "Nhttc: A gradient-based framework for generalized anticipatory collision avoidance," arXiv preprint arXiv:1907.05945, 2019.

[14] L. Wang, A. D. Ames, and M. Egerstedt, "Safety barrier certificates for collisions-free multirobot systems," IEEE Transactions on Robotics, vol. 33, no. 3, pp. 661–674, 2017

[15] J. Grover, C. Liu, and K. Sycara, "The before, during, and after of multi-robot deadlock," The International Journal of Robotics Research, p. 02783649221074718, 2016.

[16] X. Xu, P. Tabuada, J.W. Grizzle, and A. D. Ames, "Robustness of control barrier functions for safety critical control," in Proc. IFAC Conf. Anal. Des. Hybrid Syst., Oct. 2015, pp. 54–61.

[17] P. Long, T. Fan, X. Liao, W. Liu, H. Zhang, and J. Pan, "Towards optimally decentralized multi-robot collision avoidance via deep reinforcement learning," in 2018 IEEE international conference on robotics and automation (ICRA), pp. 6252–6259, IEEE, 2018.

[18] Y. F. Chen, M. Liu, M. Everett, and J. P. How, "Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning," in 2017 IEEE international conference on robotics and automation (ICRA), pp. 285–292, IEEE,2017.

[19] M. Everett, Y. F. Chen, and J. P. How, "Collision avoidance in pedestrian-rich environments with deep reinforcement learning," IEEE Access, vol. 9, pp. 10357–10377, 2021.

[20] S. Ross, G. J. Gordon, and J. A. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," 2011.

[21] S. Ross, G. J. Gordon, and J. A. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," 2011.

[22] B. D. Ziebart, A. L. Maas, J. A. Bagnell, A. K. Dey, et al., "Maximum entropy inverse reinforcement learning.," in Aaai, vol. 8, pp. 1433–1438, Chicago, IL, USA, 2008.

[23] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction. MIT Press, 2018

[24] J. Ho and S. Ermon, "Generative adversarial imitation learning," in Proceedings of the 30th International Conference on Neural Information Processing Systems, pp. 4572–4580,2016.

[25] I. Kostrikov, O. Nachum, and J. Tompson, "Imitation learning via off-policy distribution matching," 2019.

[26] R. Dadashi, L. Hussenot, M. Geist, and O. Pietquin, "Primal wasserstein imitation learning," CoRR, vol. abs/2006.04678, 2020.

[27] D. Zhou, Z. Wang, S. Bandyopadhyay, and M. Schwager, "Fast, on-line collision avoidance for dynamic vehicles using buffered voronoi cells," IEEE Robotics and Automation Letters, vol. 2, no. 2, pp. 1047–1054, 2017.

[28] Z. Zhong, M. Nejad, and E. E. Lee, "Autonomous and semiautonomous intersection management: A survey," IEEE Intelligent Transportation Systems Magazine, vol. 13, no. 2, pp. 53–70, 2020.

[29] T.-C. Au, S. Zhang, and P. Stone, "Autonomous intersection management for semi-autonomous vehicles," in The Routledge handbook of transportation, pp. 88–104, Routledge, 2015.

[30] Fu Sen F. Lin and Cheng-Han Lee, "An Improved Algorithm of Juang's Method by Matrix Operations for Finding the Shortest Path in Networks," International Journal of Computer Theory and Engineering vol. 8, no. 2, pp. 136-139, 2016.

[31] Vinay Rishiwal, Mano Yadav, K. V. Arya, "Finding Optimal Paths on Terrain Maps using Ant Colony Algorithm," International Journal of Computer Theory and Engineering vol. 2, no. 3, pp. 416-419, 2010.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:26:38 UTC from IEEE Xplore. Restrictions apply.

Physica Scripta

INCLUSIVE PUBLISHING
TRUSTED SCIENCE

PAPER

The collisional study of EIC waves in a magnetized dusty plasma with the inclusion of a DC electric field

To cite this article: Anshu *et al* 2024 *Phys. Scr.* **99** 085602

View the article online for updates and enhancements.

You may also like

- Charged dust phenomena in the near-Earth space environment
W A Scales and A Mahmoudian
- Nonlinear electrostatic waves in the auroral plasma
S V Singh, R Rubia, S Devanandhan et al.
- Nonlinear interaction between electrostatic ion cyclotron waves and ion beams in an ion beam-plasma system
M Sugawa

This content was downloaded from IP address 14.139.251.98 on 08/07/2024 at 10:41

Physica Scripta

PAPER

The collisional study of EIC waves in a magnetized dusty plasma with the inclusion of a DC electric field

RECEIVED
24 January 2024

REVISED
10 June 2024

ACCEPTED FOR PUBLICATION
25 June 2024

PUBLISHED
4 July 2024

Anshu [1], **Jyotsna Sharma** [2] **and Suresh C Sharma** [1]

[1] Plasma and Nano-Simulation Lab, Department of Applied Physics, Delhi Technological University (DTU), Bawana Road, Delhi-110042, India

[2] Department of Physics, Amity School of Applied Sciences, Amity University Haryana, Manesar, Gurugram-122051, India

E-mail: suresh321sharma@gmail.com

Keywords: ion cyclotron wave, collisional dusty plasma, kinetic theory

Supplementary material for this article is available online

Abstract

This study investigates Electrostatic Ion Cyclotron (EIC) waves and their behaviour in weakly collisional plasmas, utilizing a proposed kinetic analytical model. The findings include alterations in EIC wave dispersion characteristics due to collisions, with parameters such as dust density, collision frequency, gyro-radius, magnetic field, density ratio, and electric field influencing wave growth rate and frequency. Temperature analysis reveals that higher electron-to-ion temperature ratios lead to increased frequency and critical drift velocity, while decreasing the growth rate. In addition, the critical drift velocity is studied for the unstable mode and it is observed that the relative density ratio increases with a reduction in critical drift velocity. Electron collisions destabilize EIC waves, while ion collisions stabilize them. Furthermore, the presence of dust particles decreases the growth rate of EIC waves as dust grain density increases. These results align with observations reported in previous literature.

1. Introduction

The widespread prevalence of plasma has made its exploration indispensable. Consequently, it's unsurprising that research on 'plasmas' [1–7] spans multiple domains, including various waves [8–14] and astrophysical phenomena [15–19]. Dusty plasmas are commonly utilized in semiconductor processing devices for various applications such as etching and deposition processes [20]. Dusty plasma is present in rocket exhaust, influencing the exhaust dynamics and interactions with the surrounding environment and the findings could be relevant for understanding and modelling the plasma phenomena in rocket exhaust plumes [21]. Dusty plasmas are employed in plasma processing reactors for material processing and surface treatments and the results from this work could be applied to improve the performance and control of these plasma processing reactors [20, 21]. EIC waves have been extensively studied using kinetic and fluid theories and their existence in laboratory plasmas has also been experimentally verified [22–25] [8, 9, 26–29]. These waves have been observed across diverse environments, spanning from fusion heating experiments [30–32], and space plasmas to laboratory experiments [33–38].

The initial experimental findings of EIC oscillations in laboratory plasmas driven by applying a positive potential to a small electrode immersed in a uniform magnetic field in a single-ended Q-machine were presented by D'Angelo and Motley [39]. The oscillations that were observed were identified to be current driven EIC instability which is theoretically anticipated by Drummond and Rosenbluth [40]. The fundamental characteristics of EIC waves lie in their propagation within magnetized plasmas, where ions and electrons oscillate perpendicular to the magnetic field due to their respective gyro-radii. The excitation mechanism of EIC waves involves the interaction between ions or electrons and the plasma waves, leading to the generation and propagation of EIC waves. EIC waves are often observed in a variety of plasma environments, including laboratory experiments, space plasmas, and astrophysical systems. In space, EIC waves have been detected in planetary magnetospheres, such as Earth's magnetosphere and the magnetospheres of Uranus [41].

© 2024 IOP Publishing Ltd

Table 1. Detailing the species and their corresponding notations under consideration.

S. no.	Species	Equilibrium density	Drift Velocity	Velocity	Mass	Charge	Temperature	Collisional frequency
1.	Electrons	n_e^0	$u_{ed}\|\|\hat{z}$	v_e	m_e	$-e$	T_e	υ_e
2.	Ions	n_i^0	0	v_i	m_i	e	T_i	υ_i
3.	Dust Particles	n_d^0	0	v_d	m_d	Q_d^0	T_d	υ_d

Furthermore, EIC waves play a significant role in astrophysical environments such as in the Earth's magnetotail and in comets [42]. A well-known low-frequency field-aligned phenomenon which has received a lot of interest from the plasma physics field is the current-driven electrostatic ion cyclotron (EIC) instability [3, 43, 44]. In contrast to other current-driven instabilities, this one has a strikingly low threshold electron drift velocity [22].

The comprehension of the fundamental process that causes these ion-cyclotron instabilities in collisional plasma has been given by Chaturvedi and Kaw [45] using fluid treatment. D'Angelo [3] investigated in his studies the interaction between the plasma's charged dust particles and ion cyclotron waves. He showed that as the positive ion EIC mode's frequency gradually rises, the amount of negatively charged particles augments. In a Q-machine, the study on EIC wave was reported by Barkan *et al* [35] and it was observed that the growth rate of the wave was increased in the presence of a substantial amount of negatively charged dust particles. Chow and Rosenberg [46] concluded in their research using the kinetic theory, as the relative dust concentration augments, the critical drift velocity falls during the excitation of EIC waves.

Chaturvedi and Kaw [45] studied the collisional effects of EIC waves in a fully ionised collisional plasma and observed the stability of obliquely propagating EIC wave to current flow across the lines of the magnetic field via fluid theory. Satyanarayana *et al* [47] theoretically studied the collisional effects of EIC wave in bottomside ionosphere and it was discovered that electron collisions are unstable and are essential for causing the EIC instability to arise. Suszcynsky *et al* [38] conducted an experimental study on EIC waves in laboratory plasma in an auroral ionosphere regime and demonstrated the effects of collisions between neutral particles and waves in a Q-machine. Bharuthram *et al* [48] analyzed the impact of collisions on EIC waves in plasma using kinetic theory, resulting that the presence of collisions modifies the wave behaviour even in the absence of a transverse electric field.

In a magnetized plasma, Sharma *et al* [9] using fluid treatment examined the excitation of EIC waves. They found a strong relationship between the wave's frequency and growth rate in presence of an electric field, where the frequency rises as the growth rate lowers. The inclusion of a DC electric field can significantly affect the properties and behaviour of EIC waves, making it an important factor to consider in the analysis of these waves [49]. Koepke and Amatucci [8] studied the influence of transverse, localized, DC electric fields on the EIC instability in a Q machine. They have observed that the frequency increases with the magnetic field strength. Anshu *et al* [25] studied the EIC wave in a collisionless plasma via kinetic treatment. They observe that with an inclusion of an electric field, the dispersion relation modifies and it would further affect the frequency and growth rate of the wave.

A theoretical model has not been proposed, as far as the authors are aware to study the combined effect of collision, temperature and dust in the transverse DC electric field in a dusty magnetized plasma via kinetic theory. Therefore, the present study aims to extend the existing knowledge [9, 25] by investigating the collisional effects on the propagation of EIC waves in a magnetized dusty plasma with the inclusion of a DC electric field. This analysis is motivated by the need to understand the complex interplay between collisional processes, electric fields, and wave dynamics in such plasma environments.

This manuscript is structured as follows in the subsequent segments: section 2 presents an intricate examination of the electrostatic ion cyclotron wave, addressing the determination of growth rate, frequency, and critical drift. Section 3 involves a comparative analysis of our results with existing observational findings. Section 4 encapsulates the summary and conclusion.

2. Analytical modelling

We studied a magnetized dusty plasma consisting of multiple collisional species that is enclosed in a uniform and static magnetic field B oriented along the z-direction. The system under consideration comprises three distinct species: ions, electrons, and dust particles. A tabulated representation of their respective notations is presented in table 1.

An electrostatic ion cyclotron (EIC) wave exhibits a characteristic frequency ω and propagation vector k oriented in the y-z plane and its alignment is roughly perpendicular to the applied magnetic field. In addition, we

Table 2. Notations and their formulas.

S. no.	Notations	Terms	Formula
1.	ω_{ep}	Electron plasma frequency	$\sqrt{\frac{4\pi n_e^0 e^2}{m_e}}$
2.	ω_{ip}	Ion plasma frequency	$\sqrt{\frac{4\pi n_i^0 e^2}{m_i}}$
3.	ω_{dp}	Dust plasma frequency	$\sqrt{\frac{4\pi n_d^0 Q_d^2}{m_d}}$
4.	v_{et}	Electron thermal velocity	$\sqrt{\frac{2KT_e}{m_e}}$
5.	v_{it}	Ion thermal velocity	$\sqrt{\frac{2KT_i}{m_i}}$
6.	v_{dt}	Dust thermal velocity	$\sqrt{\frac{2KT_d}{m_d}}$

provide proposed system with a transverse dc electric field E in the direction of x. This introduction of an electric field induces $\vec{E} \times \vec{B}$ drift causing excitation of a wave and has a magnitude $v_E\left(=\frac{cE}{B}\right)$, oriented in the y-direction and c is the speed of light. We have also taken into consideration the appropriate drifting Maxwellian distribution function for the electrons which is as follows [43, 50]

$$f_e^0 = n_e^0\left(\frac{m_e}{2\pi KT_e}\right)^{3/2} \exp\left[\frac{-1}{2}\frac{m_e}{KT_e}\{v_x^2 + (v_y - v_E)^2 + (v_z - u_{ed})^2\}\right], \tag{1}$$

where $K =$ Boltzmann constant, $f_e^0 =$ Maxwellian distribution function at equilibrium, v_x, v_y and v_z represents the velocities in their respective directions.

On the other hand, by changing (n_e^0, m_e, T_e) with (n_d^0, m_d, T_d) and (n_i^0, m_i, T) and taking $u_{ed} = 0$ in equation (1), the dust (and ions) distribution function can be obtained.

In equilibrium, the quasi-neutrality condition says

$$en_i^0 = en_e^0 + Q_d^0 n_d^0 \tag{2}$$

For frequencies in the same range as the ion gyrofrequency, $\omega \approx \omega_{ic} \gg \omega_{dc}$, where $\omega_{ic}\left(=\frac{eB}{m_i c}\right)$ represents the ion cyclotron frequency and $\omega_{dc}\left(=\frac{Q_d^0 B}{m_d c}\right)$ is the dust cyclotron frequency. So, it is possible to consider the dust particles to be non-magnetized.

Applying the BGK collision model [23, 51] the following is the expression for the kinetic dispersion relation that governs the behaviour of electrostatic modes in a collisional dusty plasma with dust particles to be non-magnetized and is expressed as follows

$$\varepsilon(\omega, k) = 1 + \sum_\alpha \chi_\alpha = 0, \tag{3}$$

where χ are the susceptibilities and α refers to the electrons, ions and dust grains, respectively.

The electron, ion and dust grains susceptibilities are given as follows

$$\chi_e = \frac{2\omega_{ep}^2}{k^2 v_{et}^2}\left[1 + \frac{\omega_1 + i\upsilon_e}{k_z v_{et}}\sum_n \Gamma_n^e Z\left(\frac{\omega_1 + i\upsilon_e}{k_z v_{et}}\right)\right]\left[1 + \frac{i\upsilon_e}{k_z v_{et}}\sum_n \Gamma_n^e Z\left(\frac{\omega_1 + i\upsilon_e}{k_z v_{et}}\right)\right]^{-1}, \tag{4}$$

$$\chi_i = \frac{2\omega_{ip}^2}{k^2 v_{it}^2}\left[1 + \frac{\omega_2 + i\upsilon_i}{k_z v_{it}}\sum_n \Gamma_n^i Z\left(\frac{\omega_2 - n\omega_{ic} + i\upsilon_i}{k_z v_{it}}\right)\right]\left[1 + \frac{i\upsilon_i}{k_z v_{it}}\sum_n \Gamma_n^i Z\left(\frac{\omega_2 - n\omega_{ic} + i\upsilon_i}{k_z v_{it}}\right)\right]^{-1}, \tag{5}$$

$$\chi_d = \frac{2\omega_{dp}^2}{k^2 v_{dt}^2}\left[1 + \frac{\omega + i\upsilon_d}{k v_{dt}} Z\left(\frac{\omega + i\upsilon_d}{k v_{dt}}\right)\right]\left[1 + \frac{i\upsilon_d}{k v_{td}} Z\left(\frac{\omega + i\upsilon_d}{k v_{dt}}\right)\right]^{-1}, \tag{6}$$

where $\omega_1 = \omega - k_y v_E - k_z u_{ed}$ and $\omega_2 = \omega - k_y v_E$; k_y and k_z are the wave vector oriented along y and z-direction; $\Gamma_n^\alpha = I_n(\mu_\alpha)\exp(-\mu_\alpha)$, where $\mu_\alpha\left(=\frac{k_y^2\rho_\alpha^2}{2}\right)$, $\rho_\alpha\left(=\sqrt{\frac{v_{\alpha t}^2}{\omega_{\alpha c}^2}}\right)$ denotes the gyroradius, I_n represents the modified nth order Bessel function; Z depicts the plasma dispersion function [52] and other parameters are tabulated in table 2.

Analytical solutions of equation (3) are possible in some regimes by introducing some simplifying conditions. We are only concerned with the phase-velocity regime for the EIC waves. (Drummond and Rosenbluth) [40]

$$\frac{\omega_1 + i\upsilon_e}{k_z v_{et}} \ll 1, \tag{7}$$

$$\frac{\omega_2 - n\omega_{ic} + i\upsilon_i}{k_z v_{it}} \gg 1, \tag{8}$$

$$\frac{\omega + i\upsilon_d}{k v_{dt}} \gg 1 \tag{9}$$

Equation (4) can be simplified by assuming $\mu_e \ll 1$, and utilized only $n = 0$ term because $\Gamma_0^e = 1$ and $\Gamma_n^e = 0$ when $n \neq 0$. This presumption allows us to conveniently expand the plasma dispersive function $Z(\xi)$ within the framework of kinetic theory [53] as follows:

$$Z(\xi_\alpha) = -2\xi_\alpha + \frac{2}{3}\xi_\alpha^3 + ...+i\sqrt{\pi}e^{-\xi_\alpha^2}, \quad \text{if } \xi_\alpha \ll 1 \tag{10}$$

$$Z(\xi_\alpha) = -\frac{1}{\xi_\alpha} - \frac{1}{2\xi_\alpha^3} + ...+i\sqrt{\pi}e^{-\xi_\alpha^2}, \quad \text{if } \xi_\alpha \gg 1, \tag{11}$$

where $\xi_e = \frac{\omega_1 + i\upsilon_e}{k_z v_{et}}$, $\xi_i = \frac{\omega_2 - n\omega_{ic} + i\upsilon_i}{k_z v_{it}}$ and $\xi_d = \frac{\omega + i\upsilon_d}{k v_{dt}}$.

So, equation (4) can be rewritten after substituting equations (7) and (10)

$$\chi_e = \frac{2\omega_{ep}^2}{k^2 v_{et}^2}\left[1 + \frac{\sqrt{\pi}\upsilon_e}{k_z v_{et}} + i\sqrt{\pi}\frac{\omega_1}{k_z v_{et}}\left(1 + \frac{\sqrt{\pi}\upsilon_e}{k_z v_{et}}\right)\right]. \tag{12}$$

For the ions, we can approximate the equation (5) by retaining only n = 0 and n = 1 terms in the χ_i summation; including ions-neutral collisions and further assuming $\omega \approx \omega_{ci}$, $\frac{\upsilon_i}{\omega} \ll 1$ and $\frac{\upsilon_i}{\omega - \omega_{ic}} \ll 1$, the equation reduces to

$$\chi_i = \frac{2\omega_{ip}^2(\omega_2 + i\upsilon_i)\left(1 - \frac{1 - \Gamma_0^i}{\mu_i} - \Gamma_1^i\right) - \omega_{ic}(1 - \Gamma_0^i)}{k^2 v_{it}^2(\omega_2 - \omega_{ic}) - i\upsilon_i\left[\frac{\omega_2 - \omega_{ic}}{\omega_2}\Gamma_0^i - (1 - \Gamma_1^i)\right]}. \tag{13}$$

Equation (6) can be solved by substituting equations (9) and (11), here we have neglected the dust collision frequency because the mass of dust is very large and hence, this equation abridges to

$$\chi_d = -\frac{\omega_{dp}^2}{\omega^2}. \tag{14}$$

In this context, we can ignore this term $\frac{\omega_{dp}^2}{\omega^2} \ll 1$. Also, $\frac{\omega_{dp}^2}{\omega^2} \ll \chi_i$.

Substituting equations (12)–(14) in equation (3), we get

$$\begin{aligned}\varepsilon(\omega, k) = 1 &+ \frac{2\omega_{ep}^2}{k^2 v_{et}^2}\left[1 + \frac{\sqrt{\pi}\upsilon_e}{k_z v_{et}} + i\sqrt{\pi}\frac{\omega_1}{k_z v_{et}}\left(1 + \frac{\sqrt{\pi}\upsilon_e}{k_z v_{et}}\right)\right] \\ &+ \frac{2\omega_{ip}^2(\omega_2 + i\upsilon_i)\left(1 - \frac{1 - \Gamma_0^i}{\mu_i} - \Gamma_1^i\right) - \omega_{ic}(1 - \Gamma_0^i)}{k^2 v_{it}^2(\omega_2 - \omega_{ic}) - i\upsilon_i\left[\frac{\omega_2 - \omega_{ic}}{\omega_2}\Gamma_0^i - (1 - \Gamma_1^i)\right]}\end{aligned} \tag{15}$$

This results in the EIC waves' dispersion relation for collisional plasma.

The wave is either growing or damped, the dispersion relation is used to determine the solutions for the EIC waves. The following equations determine the real and imaginary components of the frequency

$$\varepsilon_r(\omega, k) = 0|_{\omega=\omega_r}, \tag{16}$$

where ε_r is the real part of equation (15)

The frequency imaginary part is defined by [54]

$$\gamma = \left.\frac{-\varepsilon_i(\omega, k)}{\frac{\partial\varepsilon_r(\omega, k)}{\partial\omega}}\right|_{\omega=\omega_r}, \tag{17}$$

where ε_i denotes the imaginary part of equation (15)

Determining the real and imaginary parts of the equations (16) and (17) and leading to

$$\omega_r = \omega_{ic}(1 + \Delta) + k_y v_E, \tag{18}$$

where

$$\Delta = \frac{\Gamma_1^i}{1 - \Gamma_1^i - \dfrac{1 - \Gamma_0^i}{\mu_i} + \dfrac{T_i}{T_e}\left(1 + \dfrac{k^2 T_e}{4\pi n_e e^2} + \dfrac{1}{1 + (P_e + P_i)}\right)},$$

where $P_e = \dfrac{\sqrt{\pi}\,\upsilon_e}{k_z v_{et}}$;

$$P_i = \upsilon_i \frac{\omega_2\left(1 - \dfrac{1 - \Gamma_0^i}{\mu_i} - \Gamma_1^i\right) - \omega_{ic}(1 - \Gamma_0^i)}{(\omega_2 - \omega_{ic})^2}.$$

$$\gamma = -\sqrt{\pi}\frac{(\omega_r - k_y v_E - \omega_{ic})^2}{\Gamma_1^i \omega_{ic}}\frac{T_i}{T_e}\left[\frac{(\omega_r - k_y v_E - k_z u_{ed})}{k_z v_{et}}\left(1 + \frac{\sqrt{\pi}\,\upsilon_e}{k_z v_{et}}\right)\right] - \upsilon_i[1 - \Gamma_1^i]. \tag{19}$$

The analytical expressions for the real frequency and growth rate are given in equations (18) and (19) of the EIC waves in the presence of a transverse dc electric field If $v_E \to 0$ and $\upsilon_i = \upsilon_e = 0$, these mathematical formulations are similar to the results that Chow and Rosenberg reported in their prior works [46, 55]. Moreover, we replicate the expressions of Anshu *et al* [25] in the absence of collisions.

The wave growth rate, given by equation (19), changes to equation (20), if we account for the dust component.

$$\gamma = -\sqrt{\pi}\left[\frac{\dfrac{2\omega_{ep}^2}{k^2 v_{et}^2}\dfrac{(\omega_r - k_y v_E - k_z u_{ed})}{k_z v_{et}}\left(1 + \dfrac{\sqrt{\pi}\,\upsilon_e}{k_z v_{et}}\right)}{\dfrac{2\omega_{ip}^2}{k^2 v_{it}^2}\Gamma_1^i\dfrac{\omega_{ic}}{(\omega_r - k_y v_E - \omega_{ic})^2} + \dfrac{2\omega_{dp}^2}{\omega_r^3}}\right] - \upsilon_i[1 - \Gamma_1^i]. \tag{20}$$

The criterion $\gamma = 0$, which pertains to the point of marginal stability, is referred to as the critical drift velocity, u_{ed}. The expression for u_{ed} is provided as follows

$$u_{ed} = \frac{1}{k_z}\left[(\omega_r - k_y v_E) + \frac{\omega_{ic}\Gamma_1^i \upsilon_i[1 - \Gamma_1^i]}{\sqrt{\pi}(\omega_r - k_y v_E - \omega_{ic})^2}\frac{T_e}{T_i}\frac{k_z v_{et}}{\left(1 + \dfrac{\sqrt{\pi}\,\upsilon_e}{k_z v_{et}}\right)}\right]. \tag{21}$$

3. Results and discussion

In this manuscript, an analytical model is formulated for a weakly ionized collisional magnetized plasma to investigate the excitation of the EIC wave in the presence of negatively charged dust grains, incorporating an electric field through kinetic theory. The proposed model aims to examine the interaction between various plasma parameters, including gyro-radius, temperature, density etc, and the impact of collisions to understand the behaviour of the wave. The equations governing the EIC wave in section 2 are utilized to determine the frequency, growth rate, and critical drift. MATLAB is employed to generate graphs, based on the analytical solutions for real frequency and growth rate. In the present analysis, typical dusty plasma parameters of the laboratory conditions have been used [8, 35, 46, 56] i.e., plasma density of ion $n_i^0 = 10^9$ cm^{-3}, plasma density of electron $n_e^0 = (1 \times 10^9 - 0.2 \times 10^9)$ cm^{-3}, density of the dust grains $n_d^0 = (1 - 5) \times 10^4$ cm^{-3}, magnetic field B = (0.1 − 4.5) kG, ion mass $m_i = 39 \times 1.67 \times 10^{-24}$ gm, the electron mass $m_e = 9.1 \times 10^{-28}$ gm, the electronic charge e = 4.8×10^{-10}*statcoul*, the speed of light c = 3×10^{10} cm s^{-1}, ion temperature $T_i = 0.2$ eV, electron temperature $T_e = 0.2$ eV [8, 35, 46]; the charge of the dust grain $Q_d^0 = (10^3 - 10^4)e$, dust grain mass $m_d = 10^{12} \times 1.67 \times 10^{-24}$ gm, K = 1.38×10^{-16} erg/Kel is the Boltzmann constant , $k_z = 0.1$ cm^{-1}, $k_y = 1$ cm^{-1} [25]; the relative density ratio varies from 1–10 [8], the gyro-radius parameter varies from 0 − 3.5 [8], Electric field E = 0.35 *stat V* cm^{-1} [25], an ion collisional frequency $\upsilon_i = 1 \times 10^3$ s^{-1} [38], an electron collisional frequency $\upsilon_e = 4 \times 10^5$ s^{-1} [38, 56], the critical electron drift velocity $u_{ed} = 6 \times 10^7$ cm s^{-1} [8].

Figure 1 portrays the plot of normalized growth rate as a function of the gyro-radius parameter for varying δ i.e., (a) $\delta = 0.4$, (b) $\delta = 1.4$, (c) $\delta = 3$, and (d) $\delta = 4$, respectively. From figure 1, observing the trend, it becomes apparent that as the gyro-radius parameter values increase, the growth rate of the instability initially increases, reaching a maximum value i.e., $\gamma = 2.19 \times 10^4$ s^{-1} for $\mu_i = 0.96$ at $\delta = 4$. Following this peak, the

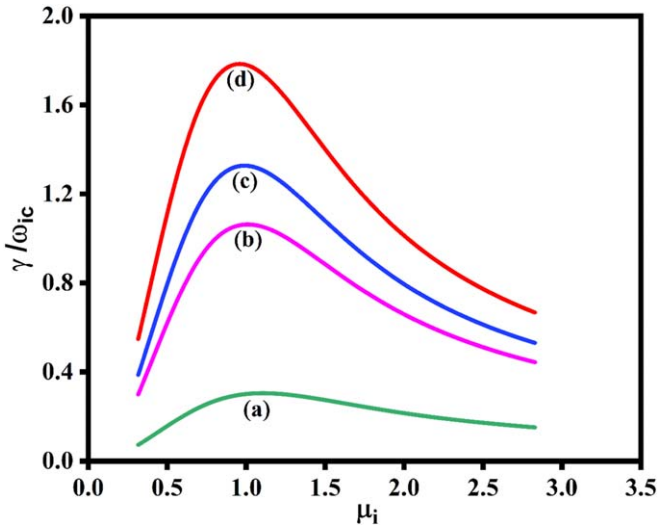


Figure 1. Variation in normalized growth rate γ/ω_{ic} with the finite gyro-radius parameter μ_i of EIC wave for varying δ i.e., (a) $\delta = 0.4$, (b) $\delta = 1.4$, (c) $\delta = 3$, and (d) $\delta = 4$..

growth rate of the instability subsequently decreases. It can also be inferred that with the increase in the relative density ratio, the growth rate of the unstable wave augments. When $\upsilon_e = \upsilon_i \rightarrow 0$ in equation (20), the growth rate increases and its corresponding value $\gamma = 2.5 \times 10^5$ s^{-1} for $\delta = 4$. This result follows the same trend as observed by the experimental observation of Chow and Rosenberg in the absence of collisions [46]. Moreover, it can also be observed that the maximum normalized growth rate of EIC wave for $\delta = 4$ is ~1.8 in the present model, whereas in our previous collisionless study, it was ~5 [25]. The reduction in the value of growth rate was observed after the inclusion of collisions can be ascribed to the increased complexity that is incorporated into the system. Because of these extra complexities, charged particles become less responsive to external perturbations, which results in a reduction in the overall growth rate.

The normalized real frequency variation corresponding to the gyro-radius parameter has been plotted in figures 2(a) and (b) for different sets of relative density ratios i.e., δ (a) $\delta = 0.4$, (b) $\delta = 1.4$, (c) $\delta = 3$, and (d) $\delta = 4$, in the presence and absence of electric field, respectively. From the plots, it can be noted that the frequency experiences an upward trend as the gyro-radius parameters increase until reaching their maximum point. Subsequently, there is a decline in magnitude for all relative density ratio values. Furthermore, an increase in the relative density ratio results in an augmentation of the real frequency. We can conclude that after the inclusion of the electric field, the real frequency follows a similar trend as in the case of the absence of the electric field despite the rise in their values. Hence, we can say that there is a significant increase in the real frequency by a factor of ~4 as the electric field included for $\delta = 4.0$ as compared with the real frequency without the electric field in the above-mentioned graphs. This is because of the reason that the Doppler effect is caused by $\vec{E} \times \vec{B}$ drift among particles, which is induced by the occurrence of an electric field. As a result, the waves' frequency grows when a dc transverse electric field is applied. The inset of figure 2(b) shows the result of $\upsilon_e = \upsilon_i \rightarrow 0$. We can say that the frequency, in collisionless case, is more as compared to collisional plasma because the charged particles can react more swiftly to an external perturbation than in collisional plasma. . Furthermore, the normalized real frequency in collisionless case ~1.46 for $\delta = 4.0$ as shown in inset of figure 2(b), which follows the same trend as observed in experimental results of Chow and Rosenberg where the value of normalized real frequency is ~1.47 [46]. Hence our result is in good agreement with the experimental observation of Chow and Rosenberg.

For the case $\delta = 1$ (no dust particles), figures 3(a) and (b) depict the impact of variation in electron collision frequency on the real frequency and growth rate, respectively. Observing the plotted data, it is evident that both frequency and growth rate values exhibit an upward trend as the gyro-radius parameter increases until they reach a maximum. Subsequently, there is a decrease in magnitude for both parameters. In both the graphs, the dash curve (a) corresponds to the collisionless case ($\upsilon_e = \upsilon_i \rightarrow 0$) and other solid curves corresponds to the collisional case for $\upsilon_i = 1 \times 10^3$ rad s^{-1} and varying υ_e i.e. (b) $\upsilon_e = 1 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (d) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (e) $\upsilon_e = 7 \times 10^5$ rad s^{-1}. It can be seen that the real frequency and growth rate (for curve (b)) for the chosen fixed plasma parameters is nearly equal to the collisionless case i.e., curve (a). This is because ion collisions have a stabilising effect whereas electron collisions have a destabilising effect [47]. For the limit, $\frac{\upsilon_e}{k_z v_{et}} \ll 1$, the ion collisions dominate electron collisions destabilising effects. As the υ_e increases to the higher value, the electron collisions start to dominate and an enhancement of the real frequency and rate of

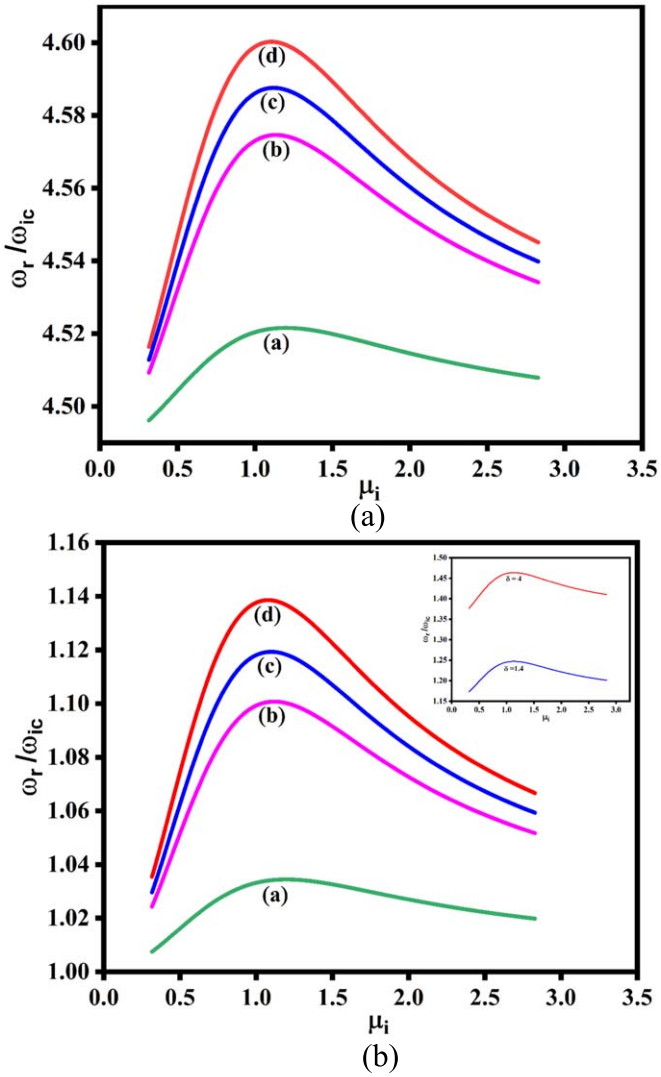


Figure 2. (a): The normalized real frequency ω_r/ω_{ic} graph with finite gyro-radius parameter μ_i of EIC wave with the electric field for varying δ i.e., (a) $\delta = 0.4$, (b) $\delta = 1.4$, (c) $\delta = 3$, and (d) $\delta = 4$. (b): Variation in normalized real frequency ω_r/ω_{ic} with the finite gyro-radius parameter μ_i without the electric field for varying δ with collisions and inset of figure 2(b) portrays the collisionless case.

growth is observed. We can infer that the wave's frequency has slightly increased while the growth rate has increased considerably when electron collision frequency augments. Our result is analogous to the Bharuthram *et al* [48] findings, where similar trend in real frequency and growth rate has been reported with an augment in the electron collisional frequencies.

The existence of negatively charged dust particles with $\delta = 3$ is the next aspect we take into account for $n_d^0 = 5 \times 10^4$ cm^{-3} in figures 4(a) and (b). Figures 4(a) and (b) depict the impact of variation of electron collision frequencies υ_e i.e. (b) $\upsilon_e = 1 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (d) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (e) $\upsilon_e = 7 \times 10^5$ rad s^{-1}. on real frequency and growth rate with gyro-radius parameter of the wave for $\upsilon_i = 1 \times 10^3$ rad s^{-1}. From the plots, it can be noted that it follows the same trend as in the above-mentioned graphs i.e., absence of dust particles with gyro-radius parameter and electron collision frequency. Moreover, we can conclude that after the addition of the dust grains in the plasma system, the growth rate and frequency for both the cases i.e., collisionless (figure 4 curve (a)) and collisional (figure 4 curve (b)–(e)) gets enhanced as compared to the previous graphs i.e., figures 3(a) and (b). Hence, we can say that after the inclusion of dust grains, growth rate upsurges by a factor of ~1.9 for $\upsilon_e = 0$ in the above-mentioned graphs. This is because the addition of the dust grains modifies the properties of the EIC wave and the collective effect of charged particles (electrons, ions and dust particles) becomes more intricate and their interaction leads to the augmentation in frequency and growth rate of the wave. Our findings matches with Chow and Rosenberg's [46] for a collisionless dusty plasma. Again, our results are in line with the findings of Bharuthram *et al* [48], because they also reported that the value of growth rate and real frequency increases after the inclusion of dust.

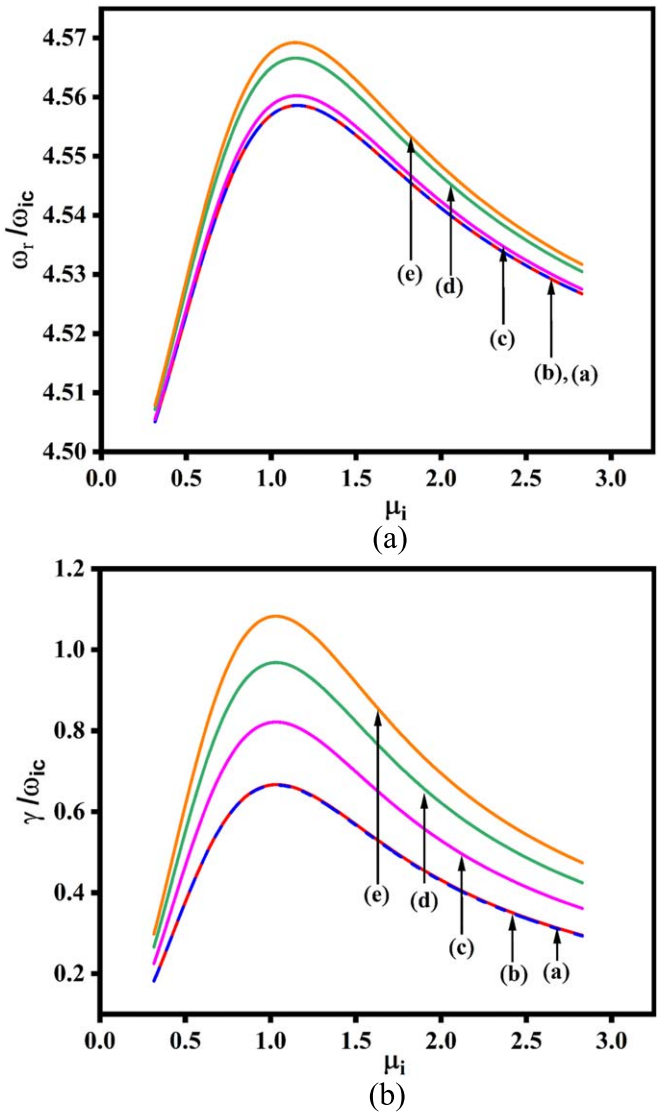


Figure 3. (a): The normalized real frequency ω_r/ω_{ic} variation for the case $\delta = 1$ corresponding to the finite gyro-radius parameter μ_i for different electron collisional frequencies υ_e i.e., (a) $\upsilon_e = 0$, (b) $\upsilon_e = 1 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (d) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (e) $\upsilon_e = 7 \times 10^5$ rad s^{-1}. (b): The normalized growth rate γ/ω_{ic} variation for the case $\delta = 1$ corresponding to the finite gyro-radius parameter μ_i for different electron collisional frequencies υ_e i.e., $\upsilon_e = 0$, (b) $\upsilon_e = 1 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (d) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (e) $\upsilon_e = 7 \times 10^5$ rad s^{-1}.

Figure 5(a) illustrates the impact of the relative density ratio on the normalized real frequency of the wave. We observed that the normalized real frequency is in a direct relationship with respect to the $\delta (=n_i^0/n_e^0)$, and it can be verified from our equation (18). Our outcome aligns with Chow and Rosenberg's theoretical conclusion [46] in the absence of an electric field and Anshu *et al* results in the absence of collisions [25].

Figure 5(b) illustrates the impact of normalized growth rate for varying υ_e i.e., (a) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (b) $\upsilon_e = 6 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 8 \times 10^5$ rad s^{-1} on the relative density ratio of the wave. As the relative density ratio increases in the plasma, the growth rate of the instability shows an augmentation. This phenomenon may be attributed to the concurrent decrease in the plasma density of electrons relative to the plasma density of ions. With an increase in the values of δ, the effective mass of ions $\mathrm{m_{ieff}}\left(=\frac{\mathrm{m_i}}{\delta}\right)$ becomes smaller than m_i, leading to enhanced mobility and, consequently, an increase in the growth rate of the instability. Figure 5(b) demonstrates unequivocally that, in accordance with the growth rate expression i.e., equation (19), the normalised growth rate increases linearly with an increase in υ_e. It is well known that electron collisions can cause the destabilization of EIC waves as shown by numerous studies [45, 47, 57]. Moreover, there is significant increase in growth rate by a factor of ~2.2 when δ changes from 1 to 4 for $\upsilon_e = 4 \times 10^5$ rad s^{-1}. Hence, our results qualitatively and quantitively both matches with the Sharma *et al* [56] findings and moreover, they have shown that, in their fluid analysis, the growth rate increases about ~2.1 when δ changes from 1 to 4 under similar conditions on the collisional EIC instability.

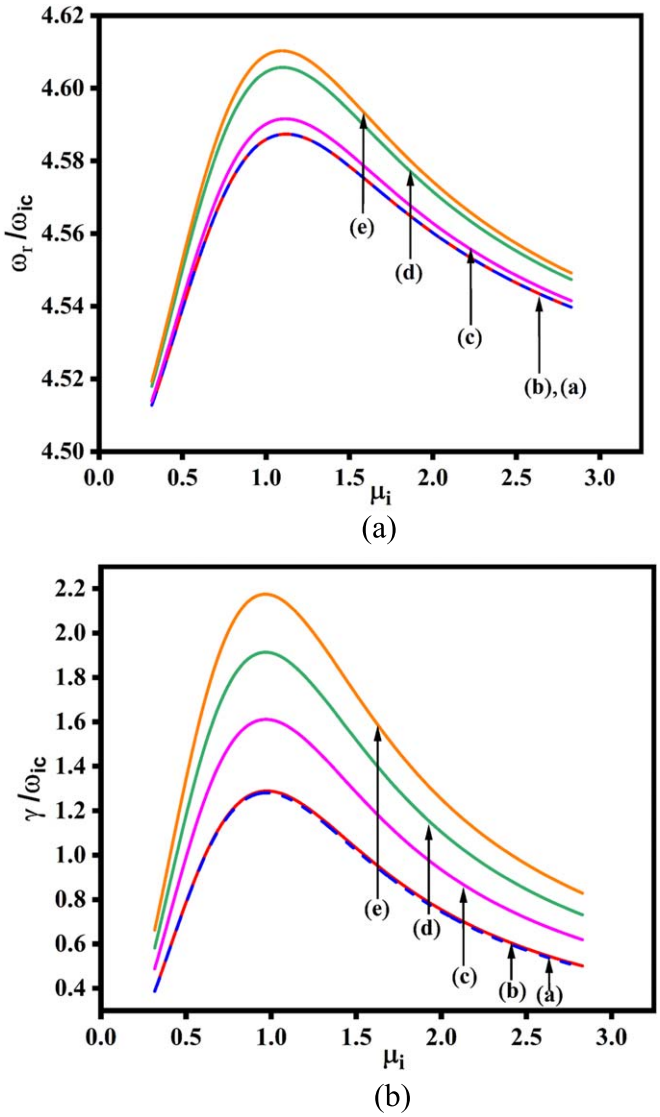


Figure 4. (a): The normalized real frequency ω_r/ω_{ic} variation for the case $\delta = 3$ corresponding to the finite gyro-radius parameter μ_i for different electron collisional frequencies i.e., (a) $\upsilon_e = 0$, (b) $\upsilon_e = 1 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (d) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (e) $\upsilon_e = 7 \times 10^5$ rad s^{-1}. (b): The normalized growth rate γ/ω_{ic} variation for the case $\delta = 3$ corresponding to the finite gyro-radius parameter μ_i for different electron collisional frequencies υ_e e.g., (a) $\upsilon_e = 0$, (b) $\upsilon_e = 1 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (d) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (e) $\upsilon_e = 7 \times 10^5$ rad s^{-1}.

For the case $\delta = 1$ (no dust particles), figures 6(a) and (b) depict the impact of variation in ion collision frequency on the real frequency and growth rate, respectively. From the graphical representations, there is an upward trend in both frequency and growth rate values as the gyro-radius parameters increase, reaching a maximum point. Subsequently, there is a reduction in magnitude for both parameters. In both the graphs, the curves correspond to the $\upsilon_e = 4 \times 10^5$ rad s^{-1} for varying υ_i i.e. (a) $\upsilon_i = 2 \times 10^3$ rad s^{-1}, (b) $\upsilon_i = 4 \times 10^3$ rad s^{-1}, (c) $\upsilon_i = 6 \times 10^3$ rad s^{-1}, (d) $\upsilon_i = 8 \times 10^3$ rad s^{-1}. It can be seen that the real frequency and growth rate reduce for increasing values of υ_i, the ion collision frequency. Further, for fixed υ_e, it is found that a rise in ion collision frequency results in the expected stabilising effect, which is also reported by Bharuthram *et al* [48]. This is because an increase in υ_i, leads to more frequent collisions between charged particles. These collisions will decrease the motion of ions reducing their ability to interact with the wave. When ions become less responsive to the wave due to increased collisions, the real frequency and growth rate of the wave decreases.

The relationship of normalized frequency and the growth rate are depicted in figures 7(a) and (b), respectively with the magnetic fields. The graphs demonstrate that the frequency and growth rate of the EIC wave correlate linearly to the magnetic field. As the magnetic field grows, so do the waves' frequency and growth rate. In the present work, the value of normalized real frequency and growth rate reaches up to ∼1.1 and ∼0.34, respectively, whereas in our previous work where collisions were not present, these values are 1.8 and 0.9, respectively [25]. It can be concluded that with the addition of collisions, the frequency and growth rate decrease

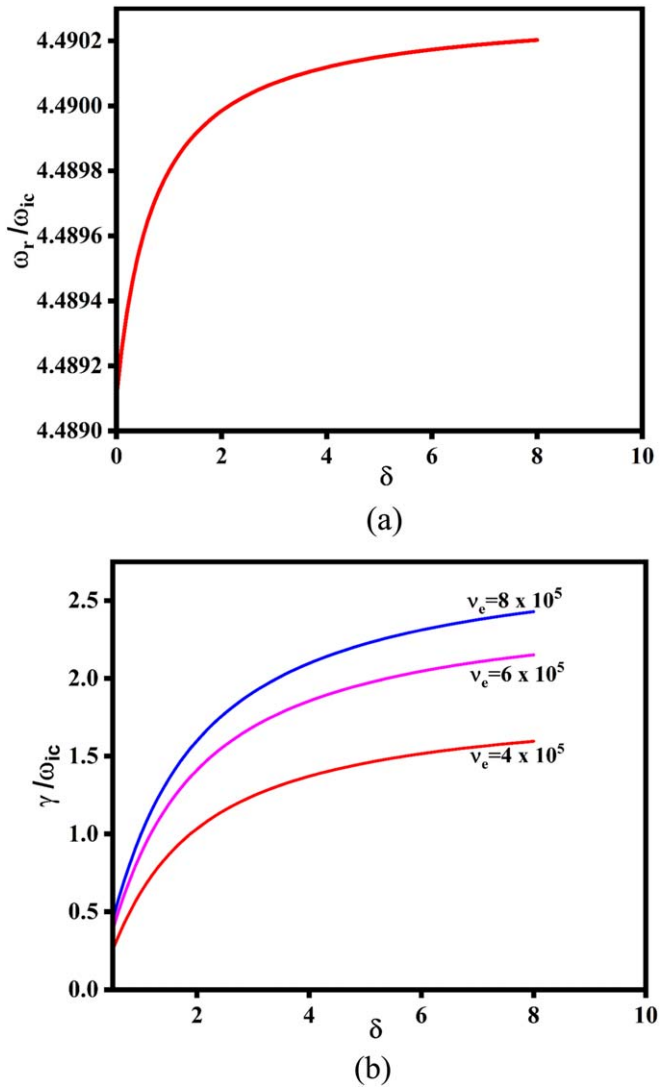


Figure 5. (a): The normalized real frequency ω_r/ω_{ic} versus relative density ratio $\delta(=n_i^0/n_e^0)$. (b): Variation in the normalized growth rate γ/ω_{ic} with relative density ratio $\delta(=n_i^0/n_e^0)$ for varying electron collisional frequencies υ_e i.e., (a) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (b) $\upsilon_e = 6 \times 10^5$ rad s^{-1} and (c) $\upsilon_e = 8 \times 10^5$ rad s^{-1}.

because the inclusion of collisions increases the complexity into the system as already discussed. The outcomes of Koepke and Amatucci [8] as well as Motley and D' Angelo [37] are comparable to this outcome without collisions. The kinetic theory yields better results than the fluid theory, according to the experimental evidence of Koepke and Amatucci [8].

In figure 8, we have displayed a normalized critical drift velocity u_{ed}/v_{et} curve with the relative density of the negatively charged dust grains. According to the graph and equation (21), the critical electron drift velocity u_{ed}/v_{et} reduces as the relative density ratio rises. Thus, as the amount of negatively charged dust particles rises, it stabilizes the EIC wave. The critical drift graph displays a trajectory that is consistent with the observed trend described in the observation done by Chow and Rosenberg [46].

As shown in figure 9, we have presented the normalized critical electron drift velocity graph for fixed $\upsilon_i = 1 \times 10^3$ rad s^{-1} for varying electron collision frequency υ_e i.e., (a) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (b) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 6 \times 10^5$ rad s^{-1}, and (d) $\upsilon_e = 8 \times 10^5$ rad s^{-1} in relation to the electron to ion temperature ratio. As we increase the T_e/T_i, critical drift rises, as seen in the graph and verified by equation (21). This is due to the fact that as the electron-to-ion temperature ratio increases, the electron temperature becomes more dominant in determining the electron mobility. With higher electron temperatures, electrons gain more kinetic energy, making them more responsive to external influences. As a result, the critical drift velocity increases as the electron-to-ion temperature ratio rises. Furthermore, with the augment in electron collision frequency, the critical drift of the EIC wave reduces [47]. This is due to the fact that numerous electron collisions might result in more effective energy dissipation and thermalization, changing the behaviour of the system as a whole and possibly stabilizing the system at lower critical electron drift velocities. We can conclude that higher

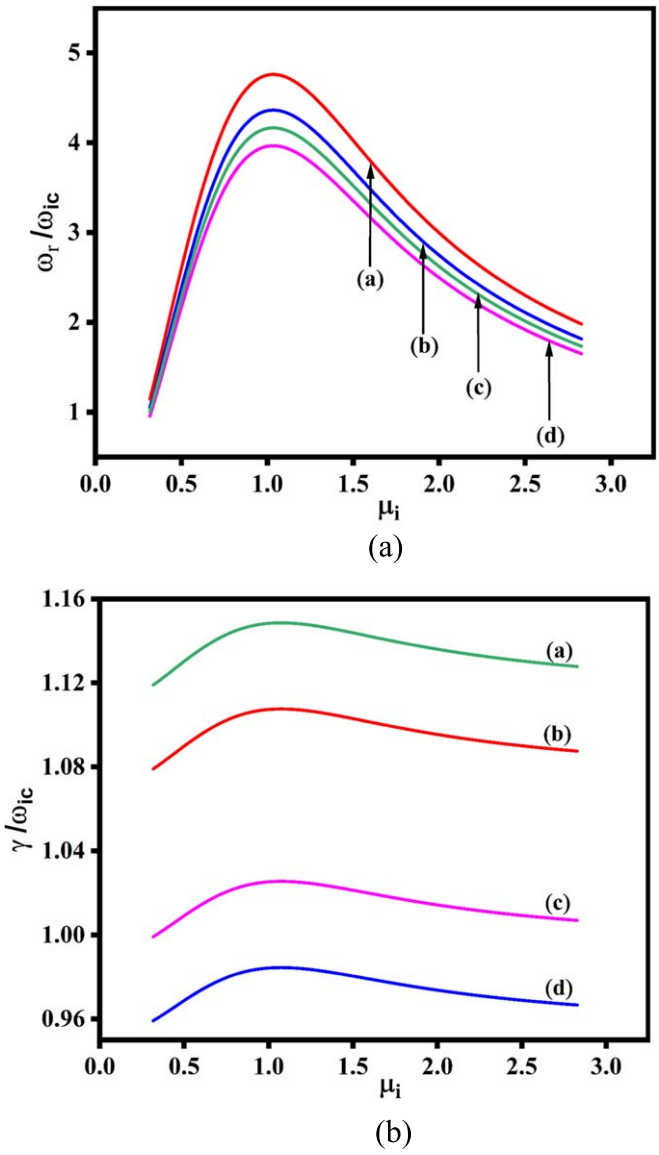


Figure 6. (a): The normalized real frequency ω_r/ω_{ic} variation for case $\delta = 1$ corresponding to the finite gyro-radius parameter μ_i for different ion collisional frequencies υ_i i.e. (a) $\upsilon_i = 2 \times 10^3$ rad s^{-1}, (b) $\upsilon_i = 4 \times 10^3$ rad s^{-1}, (c) $\upsilon_i = 6 \times 10^3$ rad s^{-1}, (d) $\upsilon_i = 8 \times 10^3$ rad s^{-1}. (b): The normalized growth rate γ/ω_{ic} variation for the case $\delta = 1$ corresponding to the finite gyro-radius parameter μ_i for different ion collisional frequencies υ_i i.e. (a) $\upsilon_i = 2 \times 10^3$ rad s^{-1}, (b) $\upsilon_i = 4 \times 10^3$ rad s^{-1}, (c) $\upsilon_i = 6 \times 10^3$ rad s^{-1}, (d) $\upsilon_i = 8 \times 10^3$ rad s^{-1}.

temperature further enhances the critical drift of the wave. Additionally, our result follows the same trend with the Levine and Kuckes [58] experimental observation in the absence of collisions.

In figure 10, the electron to ion temperature ratio in relation to the normalized growth rate graph is discussed for the EIC waves. It shows that the growth rate gradually lowers as the electron to ion temperature ratio rise and this conclusion is supported by equation (19). This is because as we have already discussed in our previous result that with increase in electron-to-ion temperature ratio, the electron mobility increases which will diminish the electron collisions there by decreasing the growth rate of the wave.

We have plotted the relationship between normalized real frequency and the electron to ion temperature ratio for varying υ_i i.e. (a) $\upsilon_i = 2 \times 10^3$ rad s^{-1}, (b) $\upsilon_i = 4 \times 10^3$ rad s^{-1}, (c) $\upsilon_i = 6 \times 10^3$ rad s^{-1}, (d) $\upsilon_i = 8 \times 10^3$ rad s^{-1} in figure 11. The normalized frequency of the wave augments with the rise in T_e/T_i. Using equation (18), it is clear that when T_e/T_i rises, Δ value rises which in turn increases the frequency. This outcome is consistent with the findings of Kindel and Kennel in the absence of collisions [23]. Moreover, higher ion collisional frequencies typically result in a reduction in EIC wave frequency.

The normalized growth rate graph versus δ for varying the values of dust number density i.e., (a) $n_d^0 = 2 \times 10^4$ cm^{-3}, (b) $n_d^0 = 4 \times 10^4$ cm^{-3}, (c) $n_d^0 = 6 \times 10^4$ cm^{-3}, and (d) $n_d^0 = 8 \times 10^4$ cm^{-3} is illustrated in figure 12. It is found that as the relative density ratio rises, the growth rate of the wave grows. According to the graphical analysis, the normalized growth rate and dust number density have an inverse

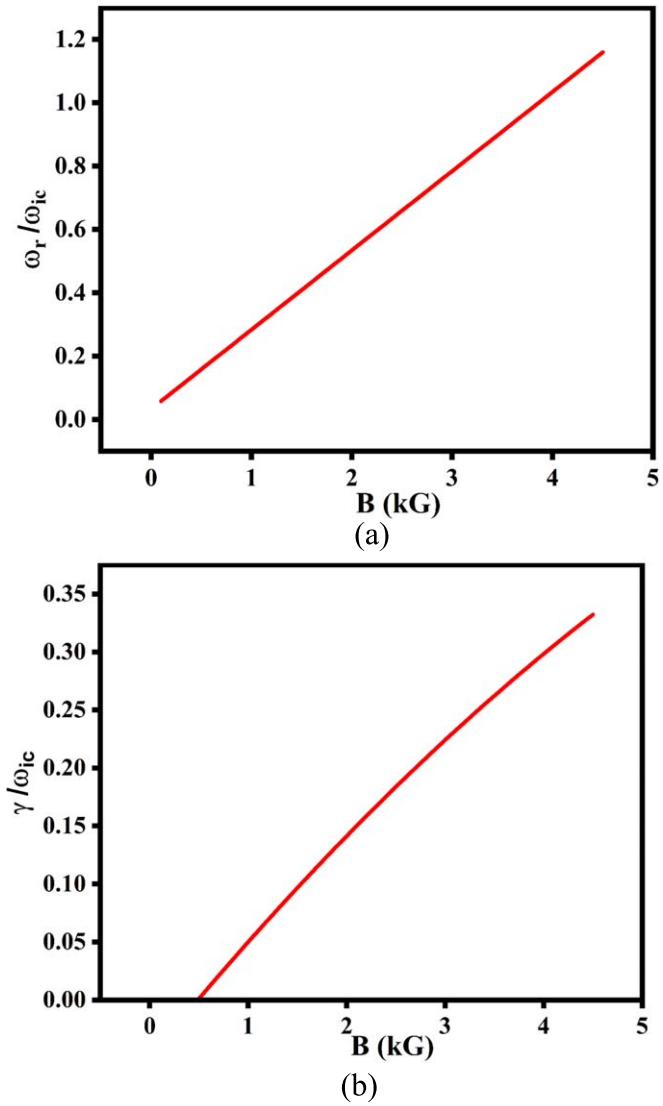


Figure 7. (a): The normalized real frequency ω_r/ω_{ic} versus magnetic field B(in kG) of EIC wave. (b): The normalized growth rate γ/ω_{ic} with respect to the magnetic field B(in kG) of the EIC wave.

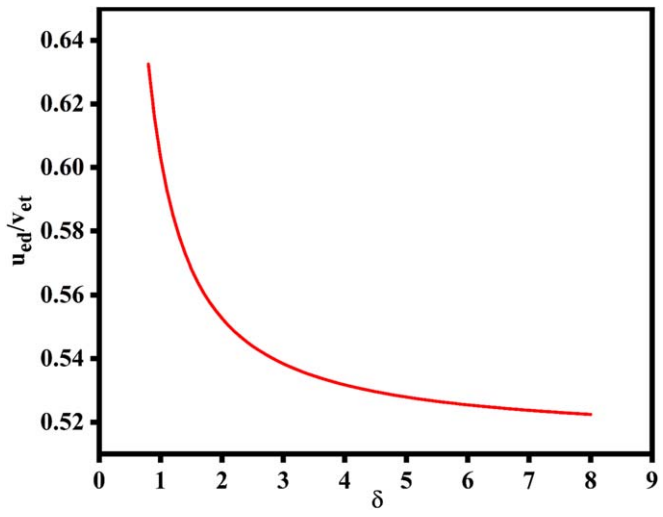


Figure 8. The normalized critical electron drift velocity u_{ed}/v_{et} in relation to the relative density of negatively charged dust particles $\delta(=n_i^0/n_e^0)$.

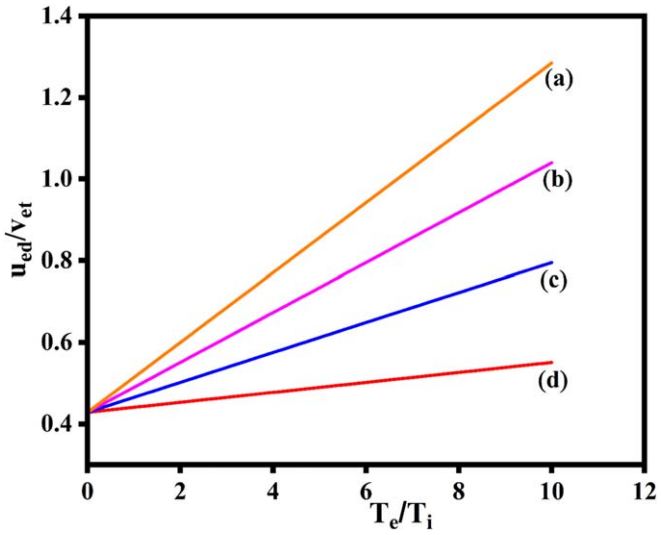


Figure 9. The normalized critical electron drift velocity u_{ed}/v_{et} for different electron collision frequencies υ_e i.e., (a) $\upsilon_e = 2 \times 10^5$ rad s^{-1}, (b) $\upsilon_e = 4 \times 10^5$ rad s^{-1}, (c) $\upsilon_e = 6 \times 10^5$ rad s^{-1}, and (d) $\upsilon_e = 8 \times 10^5$ rad s^{-1} as a function of temperature ratio (T_e/T_i).

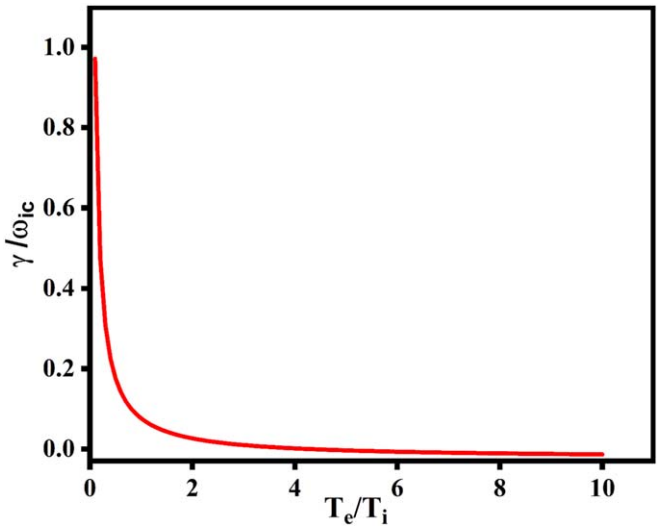


Figure 10. The normalized growth rate γ/ω_{ic} as a function of temperature ratio (T_e/T_i).

relationship, which is consistent with equation (20). This is because the rise in dust particles reduces the number of readily available electrons per dust grain. In turn, this leads to an increased need for electrons among the dust particles as a whole. This increased electron demand from the numerous dust particles leads to a fall in average dust grain charge Q_d^0, which, in turn, lowers the growth rate of the wave.

4. Conclusion

The EIC wave excitation in the weakly ionized collisional magnetized dusty plasma under the outcome of an electric field has been presented via kinetic theory. The exploration of collisional effects has been investigated across diverse plasma parameters, including growth rate, gyro-radius and frequency, among others. The wave's dispersive relationship has been deduced. The real frequency and growth rate augment with an augment in the relative density ratio. The influence of the gyro-radius parameter on both the frequency and growth rate of the wave was also assessed under varying electron and ion collision frequencies. It was observed that the frequency and growth rate initially increase with the augmentation of the gyroradius parameter before beginning to decrease as the gyroradius parameter further increases. Furthermore, it has been found that the effect of an electron collision destabilizes the mode whereas the effect of an ion collision stabilizes the mode and the growth

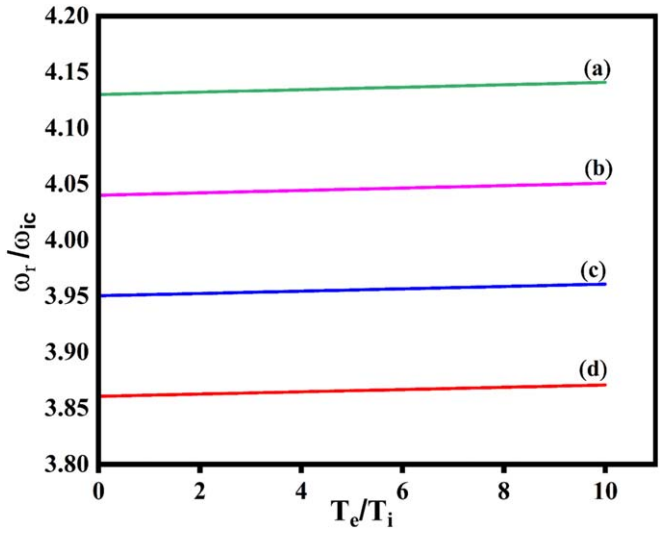


Figure 11. The normalized real frequency ω_r/ω_{ic} with varying ion collision frequency υ_i i.e. (a) $\upsilon_i = 2 \times 10^3$ rad s^{-1}, (b) $\upsilon_i = 4 \times 10^3$ rad s^{-1}, (c) $\upsilon_i = 6 \times 10^3$ rad s^{-1}, (d) $\upsilon_i = 8 \times 10^3$ rad s^{-1}. as a function of temperature ratio (T_e/T_i).

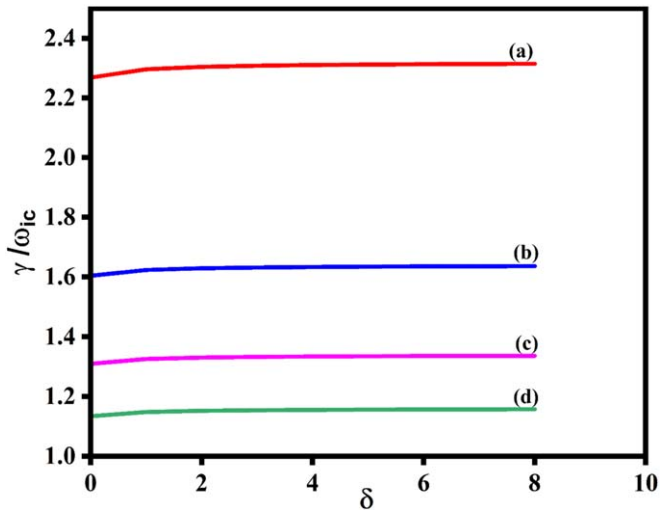


Figure 12. The normalized growth rate γ/ω_{ic} variation with $\delta(=n_i^0/n_e^0)$ for varying dust grain number density n_d^0(cm^{-3}) i.e., (a) $n_d^0 = 2 \times 10^4$ cm^{-3}, (b) $n_d^0 = 4 \times 10^4$ cm^{-3}, (c) $n_d^0 = 6 \times 10^4$ cm^{-3}, and (d) $n_d^0 = 8 \times 10^4$ cm^{-3} of the wave.

rate. The growth rate and frequency of the waves rise along with the magnetic field strength, emphasizing the stronger connection between the plasma particles and the magnetic fields. It was also discovered that as the relative density ratio rises, the critical drift reduces, which is essential for the mode's excitation. According to the study, the temperature analysis was also examined on the EIC wave and it was found that as the electron to ion temperature ratio rises, both the critical electron drift velocity and frequency rise. On the other hand, as the temperature ratio rises, the wave's growth rate reduces. It was also examined that with the increasing values of dust density, the growth rate of wave decreases. The acquired results were compared to the already existing observations [8, 23, 25, 48, 55] and are in good agreement.

The present work can be applied in laboratory plasma research [38]. This study improve our knowledge of plasma phenomena and have the potential to impact a variety of areas, including fusion energy research and could provide insights into the complex plasma dynamics in fusion devices and help in the development of more efficient and stable fusion reactors [21], tokamak and plasma heating [59, 60].

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

Conflict of interests

The authors declare no conflicts of interest.

ORCID iDs

Suresh C Sharma https://orcid.org/0000-0003-2170-195X

References

[1] Barkan A, D'Angelo N and Merlino R L 1996 Experiments on ion-acoustic waves in dusty plasmas *Planet. Space Sci.* **44** 239
[2] Bliokh P V and Yaroshenko V V 1985 Electrostatic waves in saturns rings *Soviet Astronomy* **29** 330
[3] D'Angelo N 1990 Low-frequency electrostatic waves in dusty plasmas *Planet. Space Sci.* **38** 1143
[4] Koepke M E 2002 Contributions of Q-machine experiments to understanding auroral particle acceleration processes *Phys. Plasmas* **9** 2420
[5] Shukla P K 2001 A survey of dusty plasma physics *Phys. Plasmas* **8** 1791
[6] Shukla P K and Mamun A A 2015 *Introduction to Dusty Plasma Physics* (CRC Press) (https://doi.org/10.1201/9781420034103)
[7] Anshu, Sharma J and Sharma S C 2024 Analytical modelling of inhomogeneous energy density driven (IEDD) instability in a magnetized dusty plasma cylinder *Braz. J. Phys.* **54** 8
[8] Koepke M E and Amatucci W E 1992 Electrostatic ion-cyclotron wave experiments in the WVU Q machine *IEEE Trans. Plasma Sci.* **20** 631
[9] Sharma S C, Sugawa M and Jain V K 2000 Excitation of current-driven electrostatic ion-cyclotron waves in presence of a transverse direct current electric fields in a magnetized plasma *Phys. Plasmas* **7** 457
[10] Sharma J, Sharma S C, Jain V K and Gahlot A 2013 Excitation of lower hybrid waves by a gyrating ion beam in a negative ion plasma *Phys. Plasmas* **20** 033706
[11] Sharma J and Sharma S C 2010 Excitation of electrostatic ion-cyclotron waves by an ion beam in a two-ion component plasma *Phys. Plasmas* **17** 123701
[12] Koepke M E, Carroll J J Iii and Zintl M W 1998 Excitation and propagation of electrostatic ion-cyclotron waves in plasma with structured transverse flow *Phys. Plasmas* **5** 1671
[13] Prakash V, Sharma S C, Vijayshri V and Gupta R 2014 Effect of dust grain parameters on ion beam driven ion cyclotron waves in a magnetized plasma *Progress in Electromagnetics Research M* **36** 161
[14] Verheest F 2012 *Waves in Dusty Space Plasmas* 245 (Springer Science & Business Media)
[15] Koepke M E, Carroll J J Iii and Zintl M W 1999 Laboratory simulation of broadband ELF waves in the auroral ionosphere *Journal of Geophysical Research: Space Physics* **104** 14397
[16] Chernyshov A A, Spicher A, Ilyasov A A, Miloch W J, Clausen L B N, Saito Y, Jin Y and Moen J I 2018 Studies of small-scale plasma inhomogeneities in the cusp ionosphere using sounding rocket data *Phys. Plasmas* **25** 042902
[17] Jyoti J, Sharma S C and Sharma R P 2023 Localization and turbulence of beam-driven whistler wave with magnetosonic wave in magnetopause *Phys. Plasmas* **30** 022904
[18] Gueroult R, Rax J-M and Fisch N J 2023 Wave propagation in rotating magnetised plasmas *Plasma Phys. Control. Fusion* **65** 034006
[19] Villarroel-Sepúlveda N, López R A and Moya P S 2023 The effect of heavy ions on the dispersion properties of kinetic Alfvén waves in astrophysical plasmas *A & A* **675** A84
[20] Beckers J, Berndt J, Block D, Bonitz M, Bruggeman P J, Couëdel L, Delzanno G L, Feng Y and Gopalakrishnan R 2023 Physics and applications of dusty plasmas: the Perspectives 2023 *Phys. Plasmas* **30** 120601
[21] Merlino R L 2006 Dusty plasmas and applications in space and industry *Plasma Physics Applied* **81** 73
[22] Ganguli G, Lee Y C, Chaturvedi P K, Palmadesso P J and Ossakow S L 1988 *D.C. Naval Research Lab Washington, The Effects of a DC Electric Field on the Current Driven Ion Cyclotron Instability* (Naval Research Lab)
[23] Kindel J M and Kennel cf 1971 Topside current instabilities *J. Geophys. Res.* **76** 3055
[24] Ganguli G, Lee Y C and Palmadesso P 1985 Electrostatic ion-cyclotron instability caused by a nonuniform electric field perpendicular to the external magnetic field *The Physics of fluids* **28** 761
[25] Anshu, Sharma J and Sharma S C 2022 In the existence of a transverse dc electric field, the kinetic theory of current-driven electrostatic ion cyclotron waves excitation in a magnetized dusty plasma *Contrib. Plasma Phys.* **62** e202200073
[26] Ganguli G and Palmadesso P J 1988 Electrostatic ion instabilities in the presence of parallel currents and transverse electric fields *Geophys. Res. Lett.* **15** 103
[27] Song B, Suszcynsky D, D'Angelo N and Merlino R L 1989 Electrostatic ion-cyclotron waves in a plasma with negative ions *Physics of Fluids B: Plasma Physics* **1** 2316
[28] Cattell C A, Mozer F S, Roth I, Anderson R R, Elphic R C, Lennartsson W and Ungstrup E 1991 ISEE 1 observations of electrostatic ion cyclotron waves in association with ion beams on auroral field lines from~ 2.5 to 4.5 Re *Journal of Geophysical Research: Space Physics* **96** 11421
[29] Sugai H, Kojima H and Okuda T 1982 Experimental observation of electrostatic ion cyclotron waves modified by minority ions *Phys. Lett.* A **92** 392
[30] Group T F R 1978 Ion-cyclotron instability in the TFR Tokamak *Phys. Rev. Lett.* **41** 113
[31] Waltz R E and Dominguez R R 1981 Ion cyclotron modes in tokamaks *The Physics of Fluids* **24** 1575
[32] Dumont R J and Zarzoso D 2012 Heating and current drive by ion cyclotron waves in the activated phase of ITER *Nucl. Fusion* **53** 013002
[33] Yi K-Y, Wei Z-A, Ma J X, Liu Q and Li Z-Y 2020 Excitation and characteristics of electrostatic ion-cyclotron waves launched by a grid *Phys. Plasmas* **27** 082103
[34] Cattell C, Bergmann R, Sigsbee K, Carlson C, Chaston C, Ergun R, McFadden J, Mozer F S, Temerin M and Strangeway R 1998 The association of electrostatic ion cyclotron waves, ion and electron beams and field-aligned currents: FAST observations of an auroral zone crossing near midnight *Geophys. Res. Lett.* **25** 2053

[35] Barkan A, D'Angelo N and Merlino R L 1995 Laboratory experiments on electrostatic ion cyclotron waves in a dusty plasma *Planet. Space Sci.* **43** 905
[36] Correll D L, Rynn N and Böhmer H 1975 Onset, growth, and saturation of the current-driven ion cyclotron instability *The Physics of Fluids* **18** 1800
[37] Motley R W and D'Angelo N 1963 Excitation of electrostatic plasma oscillations near the ion cyclotron frequency *The Physics of Fluids* **6** 296
[38] Suszcynsky D M, Cartier S L, Merlino R L and D'Angelo N 1986 A laboratory study of collisional electrostatic ion cyclotron waves *Journal of Geophysical Research: Space Physics* **91** 13729
[39] D'Angelo N and Motley R W 1962 Electrostatic oscillations near the ion cyclotron frequency *Physics of Fluids (New York)* **5** 633
[40] Drummond W E and Rosenbluth M N 1962 Anomalous diffusion arising from microinstabilities in a plasma *The Physics of Fluids* **5** 1507
[41] Pandey R S and Kumar M 2022 Study of electrostatic ion-cyclotron waves in magnetosphere of uranus *East European Journal of Physics* **1** 32
[42] Gomberoff L 1992 Electrostatic waves in the Earth's magnetotail and in comets, and electromagnetic instabilities in the magnetosphere and the solar wind *IEEE Trans. Plasma Sci.* **20** 843
[43] Chow V W and Rosenberg M 1996 Electrostatic ion cyclotron instabilities in negative ion plasmas *Phys. Plasmas* **3** 1202
[44] Segwal K R and Sharma S C 2018 Current-driven low-frequency electrostatic waves in a collisional strongly coupled magnetized dusty plasma *IEEE Trans. Plasma Sci.* **46** 797
[45] Chaturvedi P K and Kaw P K 1975 Current driven ion cyclotron waves in collisional plasma *Plasma Phys.* **17** 447
[46] Chow V W and Rosenberg M 1996 A note on the electrostatic ion cyclotron instability in dusty plasmas: comparison with experiment *Planet. Space Sci.* **44** 465
[47] Satyanarayana P, Chaturvedi P K, Keskinen M J, Huba J D and Ossakow S L 1985 Theory of the current-driven ion cyclotron instability in the bottomside ionosphere *Journal of Geophysical Research: Space Physics* **90** 12209
[48] Bharuthram R, Singh P, Rosenberg M and Chow V W 1998 Collisional effects on the electrostatic ion cyclotron instability in a dusty plasma *Phys. Scr.* **1998** 223
[49] Sharma S C and Hamaguchi S 2004 Current-driven dust ion-cyclotron waves in presence of a transverse DC electric fields in magnetized plasma with charge fluctuations arXiv preprint physics/0410141
[50] Rasmussen J J and Schrittwieser R W 1991 On the current-driven electrostatic ion-cyclotron instability: a review *IEEE Trans. Plasma Sci.* **19** 457
[51] Ossakow S L, Papadopoulos K, Orens J and Coffey T 1975 Parallel propagation effects on the type 1 electrojet instability *J. Geophys. Res.* **80** 141
[52] Fried B D and Conte S D 2015 *The Plasma Dispersion Function: The Hilbert Transform of the Gaussian* (Academic)
[53] Bers A, Rosenbluth M N and Sagdeev R Z 1983 Rosenbluth, Marshall N., and Roald Z. Sagdeev, eds. Handbook of Plasma Physics: Physics of Laser Plasma. Edited by A. Rubenchik and S. Witkowski. North-Holland Publishing Company, 1983.
[54] Krall N A, Trivelpiece A W and Gross R A 1973 Principles of plasma physics *Am. J. Phys.* **41** 1380
[55] Chow V W and Rosenberg M 1995 Electrostatic ion cyclotron instability in dusty plasmas *Planet. Space Sci.* **43** 613
[56] Sharma S C, Sharma K and Gahlot A 2013 Effect of dust charge fluctuations on current-driven electrostatic ion-cyclotron instability in a collisional magnetized plasma *Phys. Plasmas* **20** 053704
[57] Milić B 1972 Spontaneous excitation of the long-wave ion-cyclotron and ion-acoustic oscillations in fully ionized plasmas *The Physics of Fluids* **15** 1630
[58] Levine A M and Kuckes A F 1966 Excitation of electrostatic ion cyclotron oscillations *The Physics of Fluids* **9** 2263
[59] Sumida S, Shinohara K, Ichimura M, Bando T, Bierwage A, Kobayashi T, Yamazaki H, Moriyama S and Ide S 2023 Observation of ion-cyclotron-range-of-frequency wave emission in electron-cyclotron-resonance-heated tokamak plasma *Plasma Phys. Control. Fusion* **65** 075002
[60] Hillairet J 2023 Review on recent progress in ion cyclotron range of frequency systems, experiments and modelling for magnetic confinement fusion *Reviews of Modern Plasma Physics* **7** 16

arXiv:2407.14520v1 [physics.gen-ph] 5 Jul 2024

TILING SPACES AND THE EXPANDING UNIVERSE: BRIDGING QUANTUM MECHANICS AND COSMOLOGY

Aman Yadav[1] and Marcelo Maciel Amaral,[2]
[1]*Department of Civil Engineering, Delhi Technological University, Delhi 110042, India*
[2]*Quantum Gravity Research, Los Angeles, CA 90290, USA*
Correspondence: amanyadaav11@gmail.com

ABSTRACT

We propose a heuristic model of the universe as a growing quasicrystal projected from a higher-dimensional lattice. By extending the Schrödinger equation for a particle in a box with time-dependent boundaries, we derive an equation that resembles the Friedmann equation, addressing the Hubble tension. This model incorporates phonons and phasons, providing insights into cosmic-scale dynamics and the universe's expansion. We outline a pre-inflation tiling space phase with quantum error correction, a radiation phase dominated by quasiparticles, and a post-radiation phase with the emergence of matter. Our hypothesis, which posits that the universe is a growing quasicrystal, suggests that the necessity for an inflationary period may be obviated. Furthermore, phonon arising from this quasicrystalline structure could act as dark matter, influencing the dynamics of ordinary matter while remaining mostly undetectable by electromagnetic interactions. This hypothesis draws parallels with other crystalline matter at atomic scales that are fundamentally quantum in nature. We also explore how the notion of tiling space can support continuous symmetry atop a discrete structure, providing a novel framework for understanding the universe's expansion and underlying structure. Consequently, it is logical to start with quantum mechanics as the foundation of our model. Further development could enhance our understanding of cosmic expansion and the underlying structure of the universe.

***K*eywords** Quantum Gravity · Cosmology · Hubble Tension · Quasicrystals

Contents

1 Introduction **2**

2 From Schrödinger to Friedmann Equation **3**

3 The Universe Expansion from Quasicrystalline Tiling Spaces **7**

3.1 Universe as a Growing Quasicrystal Hypothesis 8

3.1.1 Tiling Spaces: From Discrete to Continuous 11

3.1.2 Growth Dynamics in a Quasicrystalline Universe and Dark Energy 12

3.1.3 Phonons and Dark Matter 12

3.2 Implications of the Expanding Universe as a Growing Quasicrystal 13

4 Discussion **14**

1 Introduction

Recent developments in theoretical physics have propelled us toward a deeper understanding of spacetime and gravity. The field of emergent gravity [1] suggests that spacetime and gravity might be the byproducts of fundamental quantum mechanical processes. This paradigm shift is supported by theories such as loop quantum gravity and spin foam [2], as well as superstring theory and M theory [3], all of which propose a quantum-level discreteness underlying gravity.

A key focus since the advent of perturbative superstring theory has been the quest to unravel quantum gravity [4]. This endeavor involves comprehending the theory's foundational principles and extracting insights about Planck-scale physics [5]. Consequently, there is a growing trend to incorporate concepts from condensed matter or solid-state physics, such as viewing spacetime as a deformable solid or crystal [6]. More recently, quasicrystals have been investigated [7, 8].

Clues that our theories about reality are incomplete are emerging in various fields, from anomalies in particle physics experiments to unresolved problems in cosmology [9]. For instance, the discovery of neutrino oscillations challenged the Standard Model of particle physics, indicating that neutrinos have mass. Similarly, in cosmology, the Hubble tension [10] represents a profound discrepancy between measurements of the universe's expansion rate. On one side, observations of the Cosmic Microwave Background (CMB) by missions like the European Space Agency's Planck satellite suggest a slower expansion rate, based on the state of the universe shortly after the Big Bang and theoretical models of its evolution—around 67-68 km/s/Mpc. On the other, direct measurements of the current universe, notably from the Hubble Space Telescope and confirmed by the James Webb Space Telescope, indicate a significantly faster rate [11]—around 73-74 km/s/Mpc. This tension challenges our understanding of the cosmos and suggests that our theoretical models of the universe may be incomplete, indicating that new physics might be necessary to resolve the discrepancy. Despite the increased precision of observational instruments and techniques, this persistent tension highlights a crucial gap in our understanding of fundamental cosmic processes, making it a focal point for theoretical and observational cosmology alike.

Our research adopts a dual approach within this context. First, we start with conventional quantum mechanics. While the universe's expansion is typically addressed from a classical perspective on large scales, this expansion must also extend down to microscopic scales, as suggested by various approaches to quantum cosmology [12, 13]. Unlike conventional quantum cosmology, we analyze the wave function of a particle with mass m in a specific spatial domain using the Schrödinger equation, thereby focusing on intermediate scales. Crucially, we posit that the wavenumber, wavelength, and quantum amplitudes become functions of the dynamical size of the box and are subject to the Hubble law explicitly. Considerations about the dynamic boundary of the spatial domain lead us to a version of the Friedmann equation, crucial in cosmological models, thereby underscoring the emergence of classical spacetime from quantum mechanics.

Second, drawing inspiration from the concept of a deformable solid or crystal and quantum cosmology, which deals with discreteness, we propose modeling spacetime as a growing quasicrystal. Characterized by long-range order without periodicity, this structure reflects the aperiodic patterns found in condensed matter physics. Our principal finding suggests that the cosmological scale factor might be interpreted as an inherent scaling characteristic within a quasicrystal tiling space framework. We also consider the implications of this model for the Hubble tension, offering a fresh perspective on the general evolution of the universe and providing new insights into other issues in current cosmology models, such as dark matter and dark energy.

Our research adopts a dual approach that intersects at the idea of rescaling the underlying geometry, accompanied by emerging constraints pertinent to physics at varying scales. This approach aligns with our methodology of conceptualizing spacetime through the lens of a quasicrystalline structure. By integrating scale invariance into our framework, we aim to forge a connection between the microscale realms of quantum mechanics and the macroscale phenomena observed in cosmology. This synthesis offers a comprehensive and unified perspective, potentially enhancing our understanding of the universe's nature.

In the following sections, we will further elaborate on the mathematical foundations of our model, exploring how energy considerations translate into cosmological parameters. Moreover, we will delve into the concept of spacetime as a tiling space, drawing parallels with quasiperiodic structures in condensed matter physics to bolster our hypothesis of emergent spacetime and gravity.

2 From Schrödinger to Friedmann Equation

General relativity establishes a correspondence between the geometric aspects of spacetime and the distribution of matter and energy, as encapsulated by Einstein's local field equations,

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{1}$$

where $G_{\mu\nu}$ is the Einstein tensor representing the curvature of spacetime, $g_{\mu\nu}$ is the metric tensor describing the geometry of spacetime, $T_{\mu\nu}$ is the stress–energy tensor reflecting the distribution and flow of energy and momentum in spacetime, Λ is the cosmological constant, G is the Newtonian constant of gravitation, and c is the speed of light in vacuum. On the other hand, quantum mechanics introduces the principle of wave-particle duality, which manifests both wave-like and particle-like properties in matter and energy, as illustrated by the De Broglie relations and the Schrödinger equation-Equations 2, 3. In this sense, both equations of general relativity and quantum mechanics have matter on one side and something else on the other side of an equality. This "something else" being geometry-like or wave-like. In this discourse, we explore a simplified model within the context of cosmology, where the wave aspect of quantum mechanics is proportionally linked to the geometry of spacetime. This approach specifically considers the cosmological scale factor that characterizes the rescaling of spacetime within a homogeneous and isotropic universe described by a perfect fluid.

The Friedmann equations, which are central to our understanding of the universe's expansion, are primarily derived from general relativity and frameworks that integrates classical mechanics and thermodynamics. This exploration contemplates whether analogous dynamics might emerge from applying space rescaling concepts to a local particle quantum state. The conventional understanding posits that the universe expands on large scales, particularly in the vast regions between galaxies, while gravitational forces within galaxies maintain their structural cohesion. However, this view does not necessarily preclude the possibility of spacetime expansion within matter-dominated regions, which may be challenging to observe directly.

Given that matter fundamentally are quantum phenomena and acknowledging the absence of a unified quantum gravity theory, it is plausible to infer that the cosmological expansion observed at macroscopic scales originates from the spacetime fabric at the microscopic level, possibly extending down to the Planck scale. While general relativity provides a framework for understanding how matter influences spacetime curvature, the ability of classical gravity to keep matter bound at macroscopic scales, even amidst the potential expansion of the underlying spacetime, raises intriguing questions about the emergence of matter from fundamental quantum dynamics.

To explore these ideas further, our focus shifts to examining the behavior of an existing quantum state within the context of universal expansion. By investigating how quantum states respond to changes in the scale of spacetime, we may gain new insights into the interplay between quantum mechanics and cosmology, potentially shedding light on the origin of the observed cosmological expansion and its relationship to the microscopic structure of spacetime.

Consider the time-independent Schrödinger Equation for the wave function (energy eigenfunction) $\Psi(x)$, with x representing (x, y, z), of a particle with mass m under potential $V(x)$. The equation is given by:

$$\nabla^2 \Psi = -\frac{2m(E-V)}{\hbar^2}\Psi \quad , \tag{2}$$

with E, the energy eigenvalue for the stationary state $\Psi(x) = \Psi_E(x)$.

For a particle under a potential $V(x)$, the time-independent Schrödinger equation can yield different forms of solutions depending on the relationship between the energy E of the particle and the potential V. We consider the particle in a box situation with a finite potential inside and infinite outside. The box may be taken to be spherical with a given radius. Specifically, inside the box, for a constant potential, being zero, or when $E > V$, the solutions are of the form of plane waves. Conversely, for $E < V$, the solutions differ significantly. When these solutions are substituted back into the time-independent Schrödinger equation, we obtain:

$$k^2 = \frac{2m(E-V)}{\hbar^2}, \tag{3}$$

where k is the wavenumber.

This is a stationary state analysis. We now consider a scenario where spacetime itself begins to change, specifically, rescaling with a factor $a(t)$. Consequently, the volume of a given region $\eta(t_1)$ at time t_1 will relate to the volume $\eta(t_2)$ at time t_2 by

$$\eta(t_2) = \frac{a(t_2)^3}{a(t_1)^3}\eta(t_1), \tag{4}$$

ensuring that the ratio $\eta_0 = \frac{\eta(t)}{a(t)^3}$ remains constant. Thus, we can express $\eta(t)$ as $\eta_0 a(t)^3$.

Introducing time is essential to account for the rescaling effects. The problem now should be addressed by the time-dependent Schrödinger equation, as the potential depends on the dynamic radius or volume. This problem has been studied using various approaches [14] and references therein. In this context, the energies become dynamic, and a notion of energy density in quantum mechanics arises naturally [15]. Crucially, the wavenumber, wavelength, and quantum amplitudes become functions of the dynamical size of the box, and we posit that they are subject to the Hubble law explicitly. By connecting the geometry with the wave aspect of the particle-wave duality mentioned earlier, we aim to derive Friedmann-like equations.

Revisiting equation (3), we modify it by dividing both sides by the dynamical volume η, which confines the particle in space. Equation (3) encapsulates the wave-particle duality, highlighting that the matter wavelength, $\lambda = \frac{2\pi}{k}$, is a fundamental quantum mechanical attribute that determines the probability density of locating the particle within the configuration space. Notably, a particle's wavelength, inversely related to its momentum (and thus energy), serves as a bridge to cosmological dynamics. By invoking the Hubble law, expressed as $v = H\lambda$, where $H = \frac{\dot{a}}{a}$ is the time-dependent Hubble parameter, we correlate the particle's wavelength to the cosmic scale factor $a(t)$ while keeping the energy constant. This association suggests an emergent classical kinetic energy tied to the velocity v associated to the expansion. Consequently, equation (3) is transformed into

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{K\eta}{\pi^2\hbar^2}\rho, \tag{5}$$

where $K = \frac{1}{2}mv^2$ is the kinetic energy, $\rho = E/\eta$ represents the energy density, and setting $V = 0$. Note that we are simplifying the concept of energy density here. The particle energy is time-dependent, and any particle would be subject to the same expansion that occurs everywhere. This brings a notion of a particle horizon and an energy density within each particle volume expansion. This equation bears a striking resemblance to the first Friedmann equation, suggesting that the dynamics of a quantum particle in a box with a time-dependent boundary may provide insights into the expansion of the universe. By establishing a connection between the particle's wave nature and the cosmic scale factor. To align with the usual first Friedmann equation's structure, the term $\frac{K\eta}{\pi^2\hbar^2}$ must remain constant

$$K\eta = \frac{8\pi^3\hbar^2 G}{3c^2} \tag{6}$$

with kinetic energy compensating for the changing volume and establishing a direct quantum mechanical analogy to cosmic expansion dynamics.

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3c^2}\rho, \tag{7}$$

The analysis of a relativistic particle in a box with moving walls is similar to, but more complex than, the non-relativistic case. This complexity arises whether we consider a fermionic spin-1/2 particle described by the Dirac equation or a spin-0 particle described by the Klein-Gordon equation [16]. To gain insights into the relativistic case, we can apply the same analysis using the relativistic dispersion relation derived from the energy-momentum relation:

$$(\hbar kc)^2 = E^2 - (mc^2)^2. \tag{8}$$

To bridge the quantum and classical realms, we need to determine the energy density. This involves multiplying and dividing the equation by the dynamical volume η. In the non-relativistic case, we notice the appearance of a kinetic energy term linked to the expansion. This observation suggests that the dynamical energy E, for a unique particle, can transform into different forms during the expansion process. Specifically, in the case of photons, where, $mc^2 = 0$, we consider the energy density related to the photon's energy, which evolves with the expansion. Motivated by this observation, we apply the energy-balancing constraint:

$$C = E - K - V = 0, \tag{9}$$

where E is the total energy, K is the kinetic energy, and V is the potential energy, all associated with the dynamical expansion process. This constraint allows us to examine the system's properties across different scales, connecting the microscopic energy E to macroscopic scales potentially represented by K and V. Here, K and V can be viewed as effective parameters encapsulating average or microscopic system properties as the system expands.

The energy-balancing constraint in Equation 9 provides a framework for understanding how the total energy E is partitioned into kinetic and potential components during the expansion process. By considering the conservation

of energy and the interplay between these different forms of energy, we can gain insights into the dynamics of the system across various scales. In the context of the relativistic particle in a box with moving walls, the application of this constraint may lead to the emergence of terms analogous to the energy density ρ, pressure p, and possibly the cosmological constant Λ when deriving the relativistic Friedmann-like equation. By incorporating the energy-balancing constraint into the analysis, we can explore how the quantum mechanical properties of the system, such as the wave nature of the particle and the dynamical evolution of the wavefunction, relate to the classical cosmological parameters that describe the expansion of the universe.

Isolating K while keeping E and V independent, we obtain

$$E^2 = K(E + V + V^2/K). \tag{10}$$

Since the wave number is inversely proportional to the wavelength, we can apply the Hubble law as done before. Then, Equation 8 can be rewritten as:

$$H^2 = \beta^2 \frac{K\eta}{\pi^2\hbar^2}(\rho + p) + \frac{\beta^2}{\pi^2\hbar^2}V^2 - \frac{\beta^2}{\pi^2\hbar^2}(mc^2)^2 \tag{11}$$

with $\beta = v/c$, $\rho = E/\eta$ is the energy density, and $p = V/\eta$ is the pressure. Equation 11 is analog of the Friedmann equation for relativistic matter, derived from the quantum mechanical description of a particle in a box with moving walls. The first term on the right-hand side resembles the classical Friedmann equation. The second term, proportional to V^2, may be interpreted as related to the cosmological constant Λ as it does not explicitly depend on a.

To obtain the usual contribution to radiation, we set $mc^2 = 0$. Typically, radiation energy density enters the Friedmann equation as proportional to $1/a^4$, while non-relativistic matter is proportional to $1/a^3$. Here, we derive a notion of energy density for individual massive particles and photons subject to dynamical boundary conditions related to the expansion of space, and they share a similar form. The $1/a^4$ factor for radiation arises from considering the number density of photons, which scales like matter, times the photon energy, which scales with $1/a$ due to redshift as the photon wavelength expands. The redshift here is accounted for through the Hubble law applied to the wavelength.

The essential difference between non-relativistic matter and radiation is that in one case $H^2 \propto E$, and in the other, $H^2 \propto E^2$. To achieve the same notion of energy density, our approach introduces a new term for radiation. Essentially, we achieve the same result of having different phases: one in the short wavelength regime and the other at large wavelengths at later times.

Importantly, we note that different dispersion relations will lead to different dynamics for the scale parameter, and various relations may be considered. This is in line with the usual approach where different cosmological epochs have different evolutions due to varying energy densities. To motivate our quasicrystalline model presented in the next section, let us consider quantum excitations in lattice and quasicrystalline condensed matter systems. These systems are known for their phonon quasiparticles, which emerge due to the breaking of translational symmetry. Quasicrystals, however, also host phasons—a less understood type of quasiparticle. Phasons are particularly interesting because they are linked to a reduction in dimensions; this occurs as quasicrystals are formed, at least theoretically, from the projection of higher-dimensional lattices onto lower-dimensional spaces.

The dispersion relation for phonons in lattices is similar to the relativistic one in the regime of large wavelengths, but with the speed of light replaced by the speed of sound. In this regime, the phonon dispersion relation is nearly linear, and the crystal can be treated as a continuous elastic medium, with phonons behaving like classical sound waves. However, in the regime of short wavelengths, the dispersion relation becomes nonlinear, and the microscopic details of the lattice become important. In this case, phonons are sensitive to the discrete nature of the lattice, and the nonlinear dispersion relation can lead to phonon dispersion, where phonon packets with different wavelengths propagate at different speeds, causing distortion of the waveform. The transition between these two regimes occurs when the phonon wavelength becomes comparable to the interatomic spacing, with the specific wavevector or wavelength at which this transition happens depending on the details of the crystal structure and the interatomic forces.

In quasicrystals, the situation is more complicated because they do not have a simple primitive cell like periodic crystals but rather a complex, self-similar structure. As a result, the phonon dispersion relations in quasicrystals can exhibit intricate, fractal-like patterns. Despite these differences, the linear dispersion relation for phonons in quasicrystals shares similarities with the phonon branches in periodic crystals, particularly at low frequencies and long wavelengths. We will discuss more about phonons in the next section when discussing the quasicrystalline model, where they arise as candidates to contribute in both radiation (massless phonons) and dark matter regimes (massive phonons).

Phasons in quasicrystals also share similarities with those found in relativistic physics. However, a key difference is the presence of a dispersive imaginary term in the phason dispersion relation, which becomes prominent at very large wavelengths [17, 18]. This imaginary term is related to the diffusive behavior of phasons and is absent in the phonon dispersion relation.

The presence of both phonons and phasons in quasicrystals, along with their distinct dispersion relations, may have important implications for the dynamics of the scale parameter in the quasicrystalline model of the universe. The interplay between these two types of quasiparticles, their different regimes of behavior, and their potential connection to the emergence of cosmological equations could provide new insights into the relationship between quantum mechanics and cosmology.

Using $\omega = E/\hbar$, for phasons, we have

$$v^2k^2 = \frac{E^2}{\hbar^2} + i\gamma\frac{E}{\hbar} \tag{12}$$

where γ is the phason friction coefficient. The phason is a goldstone mode that arise under a symmetry breaking mechanism [17]. In the case of lattices, the energy states are defined in a Voronoi cell. For quasicrystals, there are more fundamental cells or tiles, but with one fixed, we can consider the expansion of that cell and, as before, apply the Hubble law to the wavelength. In addition, we can use Equation 10 and multiply by the volume on both sides of Equation 12, yielding

$$H^2 = \frac{K\eta}{4\pi^2\hbar^2}(\rho + p) + \frac{1}{4\pi^2\hbar^2}V^2 + i\gamma\frac{E}{4\pi^2\hbar} \tag{13}$$

where we considered the velocity associated to phasons equal to the one of the expansion. To deal with the imaginary part, this equation can be interpreted from the energy relation where in the regime of short wavelength the phason propagation dominates with $v^2k^2 = \frac{E^2}{\hbar^2}$ leading to a Friedmann like equation similar to Equation 11. In the regime of large wavelength we have $v^2k^2 = i\gamma\frac{E}{\hbar}$ which leads to diffusion and decoupling from expansion contribution. We will come back to these two regimes in context of the quasicrystalline model of the universe in next section.

To finish this analysis, we can add the acceleration equation from usual thermodynamics considerations when one derives the Friedmann equations from Newtonian physics. For this, one adds the continuity equation

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0. \tag{14}$$

Taking the derivative in Equations (5), (11), and (13), with V constant and $\beta = 1$, we get for Equations (5) and (11):

$$2H\dot{H} = \alpha\dot{\rho}, \tag{15}$$

with $\alpha = K\eta/\pi^2\hbar^2$. For Equation (13), we have

$$2H\dot{H} = \alpha\dot{\rho} + i\gamma\frac{\dot{E}}{4\pi^2\hbar}. \tag{16}$$

Now, we can derive the acceleration term in both situations. For Equation (5), substituting the continuity equation into the derivative of the Friedmann-like equation yields (with the original potential equal to zero)

$$\frac{\ddot{a}}{a} = -\frac{\alpha}{2}(\rho + 3p). \tag{17}$$

This equation describes the acceleration of the scale factor in terms of the energy density and pressure, analogous to the second Friedmann equation in general relativity.

For Equations (11), with $m = 0$

$$\frac{\ddot{a}}{a} = -\frac{\alpha}{2}(\rho + p) + \Lambda \tag{18}$$

with $\Lambda = \frac{1}{\pi^2\hbar^2}V^2$.

For Equation (13), which includes the phason contribution, the acceleration equation becomes

$$\frac{\ddot{a}}{a} = -\frac{\alpha}{4}(\rho + p) + \frac{1}{4}\Lambda + \frac{i\gamma}{4\pi^2\hbar}(E + \frac{\dot{E}}{2H}) \tag{19}$$

The additional term involving the phason friction coefficient γ and the time derivative of the energy $\dot{E}$ modifies the acceleration of the scale factor. The imaginary nature of this term suggests that it may not significantly affect the expansion of the universe in the large wavelength regime.

These acceleration equations provide insights into the dynamics of the scale factor in the presence of matter, radiation, and phason contributions. They demonstrate how the quantum mechanical description of particles in a box with moving

walls can give rise to cosmological equations that resemble the Friedmann equations of general relativity. The inclusion of the phason term in Equation (13) highlights the potential role of quasicrystalline structures and their associated quasiparticles in the evolution of the universe.

Our analysis commenced with the time-independent Schrödinger equation for a particle confined within a potential well. We then introduced several crucial modifications to the system: a general energy-balancing constraint (Equation 9), a time-dependent boundary condition representing moving walls, and a dynamically evolving volume. These additions were aimed at establishing a connection between the quantum and classical regimes across different scales. By applying the Hubble law to the wavelength of the particle and relating it to the cosmic scale factor, we derived a Friedmann-like equation (Equation 5) from the quantum mechanical description of the system.

In the relativistic extension of our analysis, we employed the Klein-Gordon equation and the relativistic energy-momentum relation (Equation 8). This led to the emergence of a new term in the resulting Friedmann-like equation (Equation 11), which bears a resemblance to the cosmological constant term in the standard Friedmann equations of general relativity.

Furthermore, we explored the implications of different dispersion relations, particularly those associated with phonons and phasons in quasicrystalline structures. The presence of an imaginary term in the phason dispersion relation (Equation 12) led to a modified Friedmann-like equation (Equation 13) that includes a dissipative contribution arising from the phason friction coefficient. This can be interpreted as having different regimes contributions.

Bridging quantum mechanics and cosmology is complex and challenging. Our simplified model offers insights but may not capture all intricacies of this transition. A comprehensive framework incorporating quantum gravity, quantum field theory, and cosmological perturbation theory might be needed to fully understand the interplay between quantum phenomena and the universe's large-scale structure.

Nevertheless, our analysis demonstrates the potential for deriving cosmological equations from the quantum mechanical description of a particle in a box with time-dependent boundary conditions at intermediate scales. The derivation of Friedmann-like equations indicates a profound connection between the microscopic quantum world and the macroscopic evolution of the universe. This suggests that the local expansion, influenced by matter distributions, is analogous to the wave-like properties of quantum particles. The inclusion of phonon and phason contributions from quasicrystalline structures further emphasizes the potential role of condensed matter physics in understanding the nature of dark energy, dark matter, and the expansion of the universe.

3 The Universe Expansion from Quasicrystalline Tiling Spaces

Quasicrystals are characterized by several distinctive properties that make them a compelling model for the universe. These properties, as discussed in [19], include:

- **Uniformly Discrete:** For a certain radius $r > 0$, every sphere of radius r contains at most one point of the quasicrystal point set, $\triangle$.
- **Relatively Dense:** There exists a radius $R > 0$ such that each sphere of this radius encompasses at least one point of $\triangle$.
- **Locally Finite:** For any compact subset Λ in $\mathbb{R}^d$, the intersection $C = \Lambda \cap \triangle$, or cluster, is either finite or empty.
- **Finite Local Complexity (FLC):** The set $\{(t + \Lambda) \cap \triangle | t \in \mathbb{R}^d\}$, often represented by translations $\triangle - \triangle$, comprises only a finite number of clusters, modulo translations.
- **Uniform Distribution:** As proven in [19], the distribution of points within quasicrystals is uniform.

The first four properties align with a universe where matter is neither infinitely dense nor arbitrarily dispersed, and where the arrangement of matter is limited to a finite number of configurations. The fifth property, uniform distribution, supports the cosmological principle by suggesting large-scale homogeneity and isotropy in the universe, while allowing for small inhomogeneities.

A key aspect of quasicrystals with the aforementioned properties is their derivation from a specific projection from a higher-dimensional lattice. The so-called cut-and-project scheme (CPS) is defined as a 3-tuple $\mathcal{G} = (\mathbb{R}^d, G, \mathcal{L})$, where $\mathbb{R}^d$ represents a real Euclidean space, G is a locally compact abelian group (or any topological group), and $\mathcal{L}$ is a lattice in $\mathbb{R}^d \times G$. For our purposes, this scheme includes a natural projection $\pi : \mathbb{R}^d \times G \to \mathbb{R}^d$. In the context of an infinite quasicrystal, the projection typically involves restricting the points to a non-empty, relatively compact subset $K \subset G$, referred to as the window. However, in our study of a finite universe and finite quasicrystals, we project the entire mother lattice onto $\mathbb{R}^d$ part by part as explained below, thereby obviating the need to consider this window.

To illustrate our approach more concretely, let us consider the 4D Elser-Sloane quasicrystal (ESQC) derived from the E8 root lattice [20], which is the most studied 4D quasicrystal. We position a coordinate system on the E8 lattice, with the origin placed at an E8 vertex. The E8 lattice can then be visualized as concentric shells of points radiating from the origin. As noted in [21], the number of points on each E8 shell increases in a nonlinear fashion, described by the function

$$N_{E8}(n) = 240 \sum_{d|n} d^3 \tag{20}$$

where the right-hand side of the equation is the sum of the cubes of all divisors d of n, multiplied by 240. Projecting shell by shell, the E8 lattice to 4D using the ESQC CPS generates a growing 4D quasicrystal.

The projection of the initial shells of the E8 lattice, when visualized from a 3D slice perspective, reveals patterns that emerge under rescaling. See Image 1 for a visual representation.

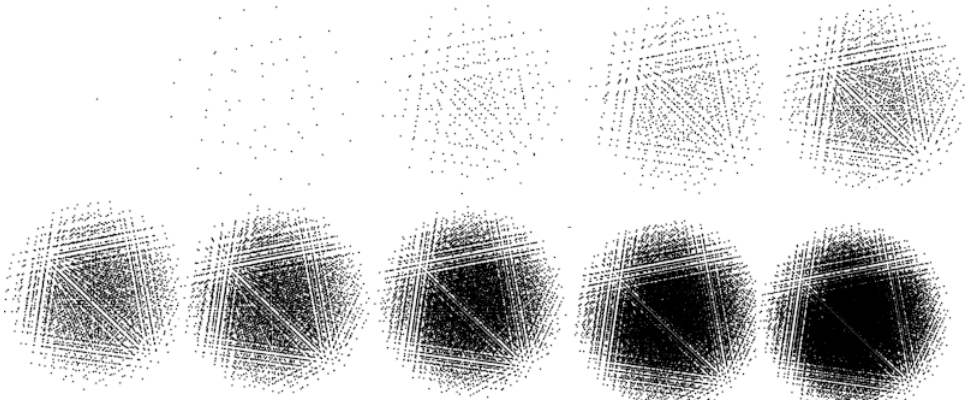


Figure 1: First 10 shells from E8 lattice to 4D ESQC to a 3D slice.

The notion of tiling spaces emerges from considering each possible tiling that can be built on a specific CPS as a point in an abstract space. In a quantum mechanical framework, each of these possible tilings can be defined as a state. However, rather than working with individual states, we employ the more comprehensive concept of Hilbert spaces, which encompass all possible states. This approach bridges the geometric nature of tilings with the probabilistic nature of quantum mechanics. Consequently, tiling spaces provide a robust framework for describing both the geometric structure and the quantum mechanical evolution of the universe during its expansion, offering a unified perspective on cosmic dynamics.

3.1 Universe as a Growing Quasicrystal Hypothesis

Our working hypothesis is that the universe is a growing quasicrystal. This is a simple heuristic model of the universe. Let us gather support for this idea and explore its consequences.

This model proposes an unconventional origin of the universe, diverging from the traditional Big Bang theory. It suggests that the universe began with the projection of a single point from a higher-dimensional lattice, such as the E8 root lattice. This initiated a sequence of point projections that sequentially constructed the universe's structure. This process, governed by a computational framework, unfolds step by step transitioning from one shell $N_{\text{E8}}(n)$ to the next $N_{\text{E8}}(n+1)$ and so forth. The number of points in each shell grows rapidly, with the first ten shells containing 240, 2160, 6720, 17520, 30240, 60480, 82560, 140400, 181680, 272160 points, respectively. The projection into a 4D quasicrystalline structure not only expands the universe but also intricately subdivides it internally, as points from newer shells intermingle with those from older shells, showcasing a complex process of growth and subdivision characteristic of the universe's early stages (see Figure 2 for a lower-dimensional illustration).

In this early stage, the universe consists of points, which are projections from the higher-dimensional root lattice. In the Lie algebra framework of modern gauge theories, these points represent fundamental particle states that form the Hilbert spaces of representations of the Lie algebra. At this stage, there is no scale or mass, and thus no distinction between particles, highlighting an analogy to the grand unification scale.

In condensed matter physics, a growing quasicrystal is associated with expanding Hilbert spaces. We can conjecture a similar process for the universe. As the universe grows, a burgeoning Hilbert space $\mathcal{H}_n$ emerges at each growth step n, reflecting the tiling spaces [22] of potential quasicrystalline configurations. To address the different tiling configurations

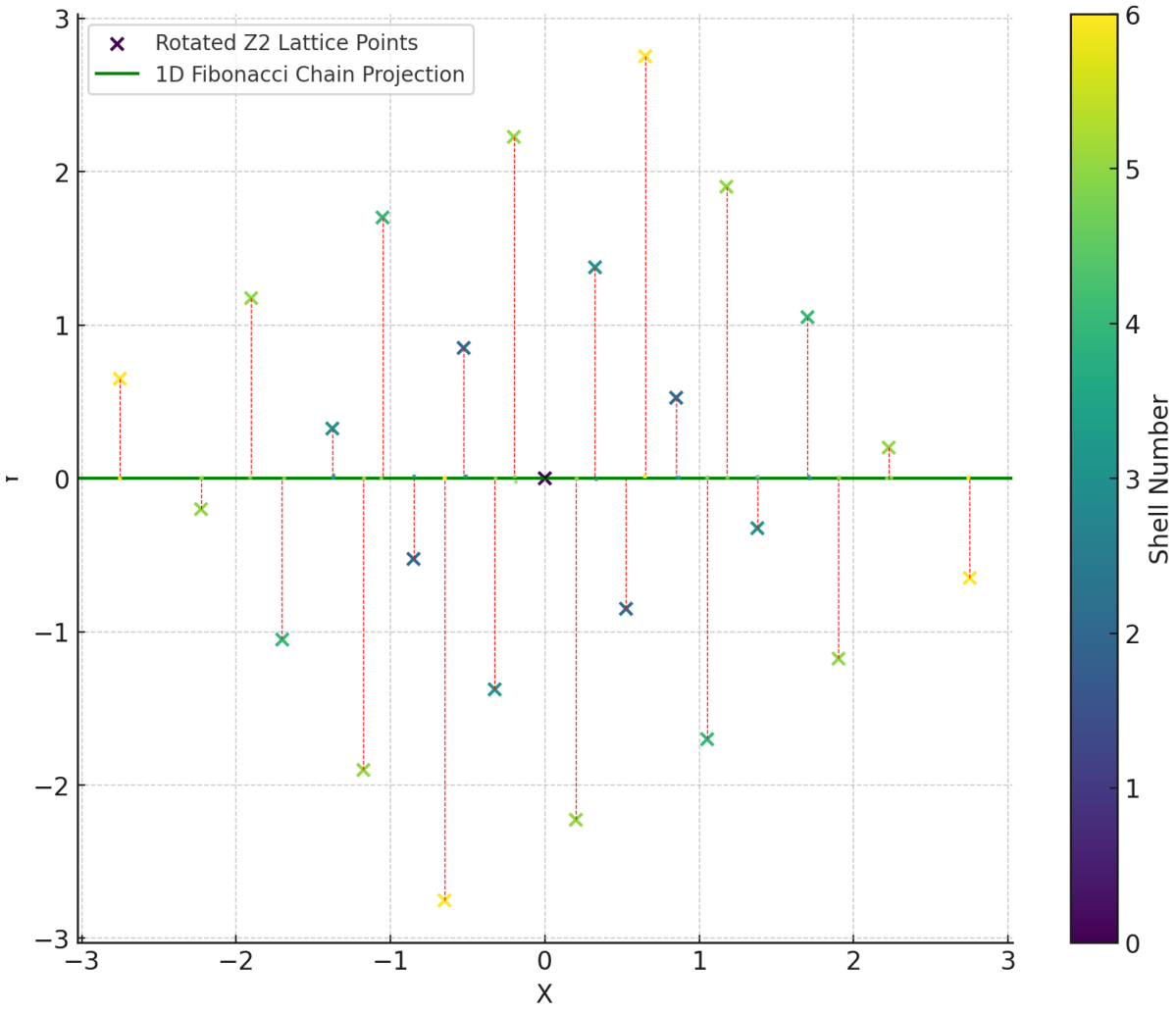


Figure 2: A rotated Z2 being projected shell by shell with the color tracking the ordering. As this one dimensional universe expands, it also get new points inside the already projected region.

of quasicrystals, it is standard to consider the associated C^*-algebra structures [23] and the notion of tiling spaces. This Hilbert space can be constructed as follows: Let $\mathcal{T}_n$ denote the set of all distinct tiling configurations at the n-th growth step, with each tiling configuration $T_n \in \mathcal{T}_n$ represented as a basis state $|T_n\rangle$ in $\mathcal{H}_n$. The growth process from step n to $n+1$ can be expressed as:

$$\mathcal{T}_n \to \mathcal{T}_{n+1}, \tag{21}$$
$$\mathcal{H}_n \to \mathcal{H}_{n+1}. \tag{22}$$

The wavefunction of the universe at the n-th growth step, $|\Psi_n\rangle \in \mathcal{H}_n$, can be written as a superposition of all possible tiling configurations:

$$|\Psi_n\rangle = \sum_{T_n \in \mathcal{T}_n} \alpha(T_n)|T_n\rangle, \tag{23}$$

where $\alpha(T_n)$ are complex coefficients determined by the specific growth rules of the quasicrystal. As the universe grows, the wavefunction evolves to incorporate the new tiling configurations that emerge at each step, $|\Psi_n\rangle \to |\Psi_{n+1}\rangle$.

This growing Hilbert space structure suggests that the early universe may be viewed as a form of a quasicrystalline topological quantum computer, with quantum error-correcting codes (QECCs) inherently embedded [24]. At each growth step, a subspace $\mathcal{C}_n \subset \mathcal{H}_n$ can be identified, spanned by a set of wavefunctions $|\Psi n, [T]\rangle$ that exhibit local

indistinguishability and recoverability properties [24]:

$$|\Psi_{n,[T]}\rangle = \sum_{g\in G_n} \beta(gT_n)|gT_n\rangle, \tag{24}$$

where $[T]$ denotes an equivalence class of tiling configurations related by the symmetry group G_n, and $\beta(gT_n)$ are complex coefficients. The subspace $\mathcal{C}_n = \text{span}(|\Psi_{n,[T]}\rangle)$ can be interpreted as the code space of a QECC at the n-th growth step, capable of correcting errors in any local region of the quasicrystal. As the universe grows, the QECC evolves, $\mathcal{C}_n \to \mathcal{C}_{n+1}$.

This growing QECC structure provides a framework for understanding how quantum information can be encoded and protected within the geometry of a quasicrystalline universe. Notably, the Hilbert spaces associated with certain quasicrystals are isomorphic to the fusion Hilbert spaces that appear in topological quantum computing systems [25]. This establishes a connection between QECCs and topological quantum computation within the context of our working hypothesis of the growing quasicrystalline universe.

The emergence of this quasicrystalline structure and its associated Hilbert spaces in the early universe may have profound implications for the evolution of spacetime, matter, and quantum information processing. As the universe continues to grow and evolve, the interplay between the quasicrystalline geometry, QECCs, and topological quantum computation may guide the formation of large-scale cosmic structures and the emergence of complex phenomena. This heuristic model offers a novel perspective on the origin and evolution of the universe, connecting ideas from condensed matter physics, quantum information theory, and cosmology.

As the quasicrystalline universe evolves, phonons and phasons should emerge naturally due to the intrinsic properties of growing quasicrystals, even those composed of 'spacetime atoms.' Phonons correspond to collective vibrational modes of the tiling configurations, while phasons are associated with the rearrangement of the quasicrystalline structure without changing the overall energy. The presence of these quasiparticles is significant, as their dispersion relations lead to a form of the Friedmann equation (Equation 13), implying the establishment of standard expansion.

To incorporate phonons and phasons into our quantum description of tiling spaces, we introduce creation and annihilation operators acting on the Hilbert space $\mathcal{H}_n$ at each growth step. Let $a^\dagger_{k,n}$ and $a_{k,n}$ denote the creation and annihilation operators for a phonon with wave vector k at the n-th growth step, satisfying the commutation relations:

$$[a_{k,n}, a^\dagger_{k',n}] = \delta_{k,k'}, \quad [a_{k,n}, a_{k',n}] = [a^\dagger_{k,n}, a^\dagger_{k',n}] = 0. \tag{25}$$

Similarly, let $b^\dagger_{q,n}$ and $b_{q,n}$ denote the creation and annihilation operators for a phason with wave vector q at the n-th growth step, satisfying similar commutation relations:

$$[b_{q,n}, b^\dagger_{q',n}] = \delta_{q,q'}, \quad [b_{q,n}, b_{q',n}] = [b^\dagger_{q,n}, b^\dagger_{q',n}] = 0. \tag{26}$$

The phonon and phason operators commute with each other:

$$[a_{k,n}, b_{q,n}] = [a_{k,n}, b^\dagger_{q,n}] = [a^\dagger_{k,n}, b_{q,n}] = [a^\dagger_{k,n}, b^\dagger_{q,n}] = 0. \tag{27}$$

Using these operators, we define phonon and phason excited states by acting on the tiling configuration states $|T_n\rangle$:

$$|T_n; \{k\}_m; \{q\}_l\rangle = a^\dagger_{k_1,n} \dots a^\dagger_{k_m,n} b^\dagger_{q_1,n} \dots b^\dagger_{q_l,n} |T_n\rangle, \tag{28}$$

where $\{k\}_m = \{k_1, \dots, k_m\}$ and $\{q\}_l = \{q_1, \dots, q_l\}$ characterize the m phonons and l phasons, respectively.

The growth of the quasicrystalline universe can be described in terms of the evolution of these excited states. As the universe grows from step n to step $n+1$, the phonon and phason creation and annihilation operators evolve:

$$a^\dagger_{k,n} \to a^\dagger_{k,n+1}, \quad a_{k,n} \to a_{k,n+1}, \tag{29}$$

$$b^\dagger_{q,n} \to b^\dagger_{q,n+1}, \quad b_{q,n} \to b_{q,n+1}. \tag{30}$$

The wavefunctions $|\Psi_n\rangle$ and $|\Psi_{n,[T]}\rangle$ can be expanded in terms of these excited states:

$$|\Psi_n\rangle = \sum_{T_n\in\mathcal{T}_n} \sum_{\{k\},\{q\}} \alpha(T_n; \{k\}, \{q\}) |T_n; \{k\}; \{q\}\rangle, \tag{31}$$

$$|\Psi_{n,[T]}\rangle = \sum_{g\in G_n} \sum_{\{k\},\{q\}} \beta(gT_n; \{k\}, \{q\}) |gT_n; \{k\}; \{q\}\rangle, \tag{32}$$

where the coefficients $\alpha(T_n; \{k\}, \{q\})$ and $\beta(gT_n; \{k\}, \{q\})$ depend on both the tiling configurations and the phonon and phason excitations.

The emergence of phonons and phasons in the growing quasicrystalline universe has important implications for the properties of the corresponding quantum error-correcting codes. The presence of these excitations introduces additional degrees of freedom that may affect the error correction capabilities of the code. As discussed in the previous section, the dispersion relations of phonons and phasons lead to Friedmann-like equations describing the expansion of the universe, contributing to the energy density and exhibiting different behaviors in various phases of cosmological evolution. In the spirit of [26], we propose that phonon and phason excitations can provide the seeds for composite matter formation. In this model, quasicrystal points act as point-particle excitations, forming composite particles such as hadrons. These excitations contribute to the radiation epoch spectrum, and as more composite matter emerges, the universe transitions to a matter-dominated epoch. As the universe expands, phonon excitations may continue to influence its evolution. In particular, the dispersion relation for phasons (Equation 12) includes an imaginary term that becomes significant at large wavelengths, leading to diffusive behavior, thereby excluding its contribution from the energy density.

3.1.1 Tiling Spaces: From Discrete to Continuous

While quantum field theory may be suitable for studying these transitions, most many-body systems, including quasicrystals and the universe, exist at finite temperatures. As a result, the equilibrium state of the system is described by a thermal density matrix, often involving mixed states. Consequently, conventional quantum field theory techniques may not be appropriate. The literature frequently employs effective field theory techniques to address these conditions. Here, we will describe tiling spaces using the probabilistic approach of [19]. Tiling spaces are defined by the continuous hull $X(\Lambda)$, which includes all translates of a point set Λ in $\mathbb{R}^d$. This framework is typically applied to infinite quasicrystals, but to understand regimes where dynamics occur at scales much larger than the underlying spacing (potentially at the Planck scale), we expect to recover invariance under continuous translations in the tiling space.

A characteristic property of a point set in $\mathbb{R}^d$ is the average number of points per unit volume, and we are interested in the relative frequency f^r of subsets or clusters of points, defined as the absolute frequency per unit volume over the density of the point set Λ. Various methods to compute this quantity can be found in the quasicrystal literature [19]. This will allow us to define a probabilistic measure in the tiling space.

These relative frequencies define an invariant probability measure on the set of tiling configurations $X_0(\Lambda)$, which consists of all tiling configurations with the origin as a point. For a non-empty finite cluster P, we define the cylinder set Z_P containing all configurations with P as a subcluster:

$$Z_P := \{\Lambda' \in X_0(\Lambda) \mid P \subset \Lambda'\}. \tag{33}$$

The measure of this cylinder set is given by the relative frequency:

$$\mu_0(Z_P) = f^r_\Lambda(P). \tag{34}$$

This measure extends to a positive measure on the entire tiling space $X(\Lambda)$ by standard arguments, using the σ-algebra generated by the cylinder sets. The invariant measureμ provides a robust framework to explore the statistical properties of quasicrystalline structures and their implications for cosmology. We can express the partition function as:

$$\begin{aligned} Z_X &= \sum_{\Lambda \in X(\Lambda)} \prod_{P \in \Lambda} f^r_\Lambda(P) = \sum_{\Lambda \in X(\Lambda)} \prod_{P \in \Lambda} \left(f^r_\Lambda\right)^{N f^r_\Lambda} \\ &= \sum_{\Lambda \in X(\Lambda)} e^{N \sum_P f^r_\Lambda(P) \ln(f^r_\Lambda(P))} = \sum_{\Lambda \in X(\Lambda)} e^{-N I_\Lambda}, \end{aligned} \tag{35}$$

where N is large, implying that the number of specific configurations P or clusters must approach its frequency. We introduce the information entropy of the tiling space as

$$I_\Lambda = -\sum_P f^r_\Lambda(P) \ln(f^r_\Lambda(P)). \tag{36}$$

To connect this framework with a notion of energy density, we start by calculating the emergent average energy $\langle E \rangle$ from the partition function Z:

$$\langle E \rangle = -\frac{\partial ln Z_X}{\partial \beta}, \tag{37}$$

where $\beta = N = \frac{1}{kT}$ is the inverse temperature. Note that N is playing the role of β. As the tiling space grows the temperature lowers.

Given the volume η, the energy density ρ is then:

$$\rho = \frac{\langle E \rangle}{\eta}. \tag{38}$$

This framework allows us to analyze the statistical properties of quasicrystalline structures and their implications for cosmology. The information entropy I_Λ plays a crucial role in understanding the distribution and frequency of clusters within the tiling space, and the derived energy density connects this microscopic structure to macroscopic physical properties at large scales. With matter distributions, there are associated dispersion relations, which can lead to dynamics as we pointed out earlier.

Another aspect to consider is the self-similarity of quasicrystals. Quasicrystal point sets and associated tilings, while not retaining the discrete translational symmetry of the lattice in the projected space, exhibit new discrete symmetries such as scaling symmetry (self-similarity). This implies that the frequencies in a growing quasicrystal are dynamic and may oscillate.

Let us consider some suggestive interpretations of our working hypothesis of the universe as a growing quasicrystal.

3.1.2 Growth Dynamics in a Quasicrystalline Universe and Dark Energy

This model incorporates two distinct types of growth, each playing a crucial role in the universe's evolution:

External Growth by Shell Projection: The first type of growth occurs through the projection of new points from a higher-dimensional lattice onto our observable universe. This process can be visualized as the universe growing outward, shell by shell, from a central point. Each new shell introduces additional points, expanding the universe's structure.

Internal Growth through Subdivision and Rescaling: The second type of growth is an internal process triggered by the introduction of matter, which naturally imposes a scale on the system. This scale can be aligned with the Planck scale, introducing minimum length and time scales. As new points are projected into the universe, they subdivide existing structures. To maintain consistency with the matter-imposed scale, the universe must undergo periodic rescaling.

To illustrate these growth processes, let's consider a simplified one-dimensional universe, the usual Fibonacci chain projected point by point:

Initial state: The universe begins with a single point and two tile lengths: short and long.

Second point: A new point arises. The original short tile becomes long, and a new short tile is introduced with the old long becoming a new long and a short.

Fourth point: With the addition of the fourth point, we now have three types of tiles. This illustrates the dynamic nature of the frequency distribution due to the asynchronicity between shell projections and the self-similarity requirements.

Fifth point: At this stage, we return to having two tile types, but at a different scale from the initial state.

If matter is defined based on the initial scale, the universe must rescale this new configuration to maintain consistency. The new long tile is rescaled to match the original long tile, and the same is done for the short tile. This rescaling process is crucial for maintaining the consistency of matter within the growing quasicrystal structure. It allows for the preservation of local minimum distances between points if the rescaling is precisely timed after a certain number of points is projected.

Matter and observers are ultimately composed of points that are continuously replaced by new points projected from a higher dimension. One analogy is the process in which a person's atoms are replaced many times throughout their life. This mechanism for dark energy can be modeled by assuming an energy constraint, leading to Λ in Equations 18 and 19.

The two distinct expansion processes may unfold differently in the early radiation-dominated epoch and the later matter-dominated epoch: Growth due to the construction of the quasicrystal shell by shell from projections from a higher dimension. Rescaling of internal subdivisions from points entering the quasicrystal from the projections, necessitated by matter fixing a scale. This dual growth mechanism qualitatively addresses the Hubble tension.

3.1.3 Phonons and Dark Matter

In our quasicrystalline universe model, we can extend the analogy with solid-state physics to propose a novel explanation for dark matter. This approach leverages the concept of phonons—quantized modes of vibrations in a quasicrystal structure—and applies it to the spacetime quasicrystal of our model.

In this framework, spacetime itself is viewed as a quasicrystal, and phonons represent vibrational modes within this spacetime quasicrystal. These phonons, particularly those with long wavelengths that persist to later times,

could potentially contribute to the dark matter component of the universe. This concept draws a parallel with how phonons interact with electrons in solid-state physics. In solid-state systems, phonons can scatter electrons, influencing phenomena such as electrical resistance and superconductivity. The interaction Hamiltonian in solid-state physics typically takes the form:

$$H_{\text{int}} = g \sum_{q,k} (a_q + a_q^\dagger) c_{k+q}^\dagger c_k, \tag{39}$$

where a_q and $a_q^\dagger$ are phonon annihilation and creation operators, respectively, and c_k and $c_{k+q}^\dagger$ are electron operators.

Extending this concept to our cosmological model, we can propose an interaction Lagrangian that couples spacetime phonons with ordinary matter (e.g., electrons):

$$\mathcal{L}_{\text{int}} = g_{ph} \sum_{q,k} (\phi_q + \phi_q^\dagger) \psi_{k+q}^\dagger \psi_k, \tag{40}$$

where ϕ_q and $\phi_q^\dagger$ are the phonon fields, and ψ_k and $\psi_{k+q}^\dagger$ are ordinary matter fields like electrons.

Dark Matter Candidate: These spacetime phonons, if weakly interacting, could serve as candidates for dark matter. They would influence the dynamics of ordinary matter while remaining largely undetectable through electromagnetic interactions. This is analogous to models where dark matter interacts with ordinary matter through new forces or mediators, modifying the dynamics of both dark matter and ordinary matter. For instance, similar to dark photon interactions in some models, phonons in the spacetime quasicrystal could potentially scatter off nuclei, producing recoil signals that might be detectable in experiments designed to look for Weakly Interacting Massive Particles (WIMPs). Furthermore, the decay or annihilation of these phonons might produce standard model particles that could be detected by telescopes or neutrino detectors. It's important to note that the quantitative details of this phonon-based dark matter model are still a work in progress. However, this approach offers a novel perspective on dark matter, grounded in the quasicrystalline structure of spacetime proposed by our model. This phonon-based dark matter hypothesis not only provides a potential explanation for the nature of dark matter but also further illustrates the deep connections between solid-state physics and cosmology suggested by our quasicrystalline universe model.

3.2 Implications of the Expanding Universe as a Growing Quasicrystal

The hypothesis that the expanding universe is a growing quasicrystal has several significant implications for our understanding of cosmology:

1. **Elimination of the Inflation Scenario:** The necessity for an inflationary period may be obviated. The large number of tiles or points, N, required for uniform distribution is achieved by construction through the imposition of a quasicrystalline structure. This uniformity, intrinsic to quasicrystals, provides a natural explanation for the observed large-scale homogeneity of the universe.

2. **Radiation Epoch:** Following the initial QECC construction phase, the universe enters a radiation-dominated epoch characterized by the emergence of phasons and phonons. These quasiparticles are analogous to relativistic matter and radiation, contributing to the early universe's dynamics. The presence of these quasiparticles can lead to unique dispersion relations influencing the energy density and expansion rate.

3. **Phason Contribution at Later Times:** At later times, as the number of points N increases, phasons do not significantly contribute to the energy density. This is due to the dominance of a diffusive imaginary term in their dispersion relation at large wavelengths, effectively damping their influence on the universe's expansion.

4. **Matter and Long Wavelength Phonons:** Matter, along with long-wavelength massive phonons, continues to play a critical role in the universe's evolution. These components contribute to the overall energy density and can interact with ordinary matter, potentially acting as candidates for dark matter.

5. **Accelerated Expansion:** The universe's accelerated expansion can be explained by the internal rescaling mechanism driven by a fixed matter scale. This rescaling process results from the continuous projection and replacement of points within the quasicrystalline structure, analogous to the cosmological constant or dark energy driving the observed acceleration in the universe's expansion.

6. **States and Symmetries:** It is, however, questionable whether one can find exactly invariant states in a discrete space-time setting [12]. In the context of our quasicrystalline model, this suggests that the emergent large-scale symmetry is an approximation resulting from the statistical mechanics framework applied to the tiling space microstates. This leads to a partition function and energy densities that, in turn, give rise to Friedmann-like equations by imposing the Hubble law on the wavelength in the dispersion relation of propagating degrees of freedom.

Overall, this model provides a novel framework for understanding the universe's structure and dynamics, integrating concepts from condensed matter physics and cosmology to offer fresh perspectives on longstanding issues such as the Hubble tension and the nature of dark matter and dark energy.

—

At this juncture, readers may question why projections occur, how the system determines which shells to project, and the nature of the computational framework underlying this higher dimensional lattice. They may also wonder if more lattices from higher dimensions are involved. These inquiries parallel conventional physics questions about the Big Bang, conservation laws, and the mechanisms underlying physical laws. Despite these profound questions, the hypothesis of a higher-dimensional framework introduces testable consequences for the cosmos—namely, the emergence of a quasicrystalline structure distinct from both continuous and traditional crystalline constructs.

If the cosmos possesses a quasicrystalline structure, this feature could potentially be identified using techniques similar to those employed in materials science for detecting crystalline arrangements. The structural nature of a material—whether crystalline, quasicrystalline, or amorphous—is revealed through distinctive diffraction patterns. In the lab, such patterns are commonly produced through electron diffraction, where an electron beam interacts with the material. The resultant diffraction pattern, which provides detailed information on atomic spacing and structure, is then recorded. The presence of definitive Bragg peaks within these patterns indicates constructive interference from regular atomic planes, revealing the material's symmetry and structure.

Applying this concept to a cosmic scale, one might consider the cosmic microwave background (CMB) radiation as a macroscopic "diffraction pattern" reflecting the universe's structure. The CMB, a relic radiation from the Big Bang, offers a primordial snapshot of the universe from 380,000 years after its inception, containing valuable structural and compositional data from that era. Detailed examination of temperature and polarization fluctuations within the CMB, as captured by observatories like the Planck satellite, allows researchers to piece together the universe's large-scale structure.

In a theoretical model positing quasicrystalline symmetry in the universe, this analysis would involve searching the CMB data for Bragg-like peaks—specific intensity points revealing an ordered pattern characteristic of quasicrystalline structures. Identifying such patterns would profoundly transform our understanding of the cosmic expanse and bridge celestial phenomena with quantum-scale observations in condensed matter physics.

4 Discussion

In this paper, we have expanded the interpretation of the wave function beyond its traditional role in estimating a particle's position probability. By exploring the Schrödinger equation for a particle in a box with time-dependent boundaries and introducing a constraint that bridges microscale and macroscale quantities, we have derived an equation that bears a structural resemblance to the Friedmann equation. This development could be instrumental in addressing pivotal cosmological challenges, such as the Hubble tension. This tension arises due to observed discrepancies in the measurement of the universe's expansion rate, as denoted by the Hubble constant H_0. Notably, H_0 shows differing values when measured in the early universe—particularly via the CMB, indicating a slower expansion rate—compared to measurements from the late universe using supernovae and Cepheid variables, which suggest a faster expansion rate. Such variations point to a potential gap in our current understanding of the universe's expansion trajectory.

Furthermore, we have hypothesized a novel heuristic model of the expanding universe conceptualized as a quasicrystalline projection from a higher-dimensional lattice. The inherent properties of quasicrystals, which resonate with cosmological principles, offer promising insights into the dynamics of the cosmic scale parameter. Our approach introduces a fresh perspective on comprehending the universe's structure and its expansion, with a specific focus on unraveling the Hubble tension, dark energy, and dark matter. In our model, the universe's scale parameter is envisioned as a function of both the non-linear increase in points from projections from a higher-dimensional lattice and the rescaling of spaces inside the matter content of the universe with new points coming to the projected tiling space. The model provides a phase pre-inflation with quantum error correction, bypassing the inflation phase, and transitioning to a phase dominated by propagating phasons, phonons, and radiation. Subsequently, it evolves to a phase characterized by the emergence of matter in the limit of very large number of tiles, culminating in a later phase with two kinds of expansion engines due to the projections. Long-wavelength massive phonons arise as candidates for dark matter.

A potential critique of this model could be the apparent contradiction between the continuous symmetry observed in the universe and the discrete nature of a quasicrystalline structure. This critique can be addressed by noting that quasicrystals give rise to tiling spaces that exhibit continuous symmetry on very large scales. We introduced a statistical mechanics framework to show how this can occur, using a probabilistic measure for tiling space microstates, which

leads to a partition function and energy densities. The Friedmann-like equations arise by imposing the Hubble law on the wavelength in the dispersion relation of propagating degrees of freedom.

To fully realize the potential of this approach, further detailed analysis is required. This will involve rigorously developing the quantitative model and enhancing its predictive power, thereby contributing significantly to our understanding of the universe's expansion and its underlying code structure.

References

[1] Verlinde, E. On the origin of gravity and the laws of Newton. J. High Energ. Phys. 2011, 29 (2011).

[2] Rovelli, C., Vidotto, F.: Covariant Loop Quantum Gravity. Cambridge University Press 1 edition, (2014).

[3] Green, M. B., Schwarz, J. H., Witten, E.: Superstring Theory, Vol. I and Vol. II, Cambridge University Press (1988).

[4] Kiefer, C. Quantum Gravity: Third Edition. (OUP Oxford,2012)

[5] Rovelli, C. Loop Quantum Gravity. Living Rev. Relativ. 1, 1 (1998). arXiv:gr-qc/9710008.

[6] Tahim, M. O., Landim, R. R., Almeida, C. A. S. Spacetime as a deformable solid. Modern Physics Letters A. Vol. 24, No. 15, pp. 1209-1217 (2009). arXiv:0705.4120 [gr-qc].

[7] M. Amaral, R. Clawson and K. Irwin, Quasicrystalline Spin Foam with Matter: Definitions and Examples, [arXiv:2306.01964 [gr-qc]].

[8] Amaral, M.; Fang, F.; Aschheim, R.; Irwin, K. On the Emergence of Spacetime and Matter from Model Sets. Preprints 2021, 2021110359. https://doi.org/10.20944/preprints202111.0359.v2

[9] Gaspar, Y., Tambor, P. The Laws of Nature and the Problems of Modern Cosmology. Found Sci (2023). https://doi.org/10.1007/s10699-023-09904-1

[10] Huterer, D. Hubble tension. Eur. Phys. J. Plus 138, 1004 (2023). https://doi.org/10.1140/epjp/s13360-023-04591-0

[11] Riess, A., & Anand, G. (2024). Webb and Hubble Affirm Universe's Expansion Rate, Puzzle Persists. The Astrophysical Journal Letters.

[12] Martin Bojowald 2015 Rep. Prog. Phys. 78 023901 DOI 10.1088/0034-4885/78/2/023901

[13] Kuzmichev, V.E., Kuzmichev, V.V. The Hubble tension from the standpoint of quantum cosmology. Eur. Phys. J. C 84, 121 (2024). https://doi.org/10.1140/epjc/s10052-024-12490-6

[14] K. Cooney. The infinite potential well with moving walls. arXiv:1703.05282 [quant-ph], (2017).

[15] Francisco Ricardo Torres Arvizu et al, On the energy density in quantum mechanics. Phys. Scr. 98, 125015, (2023).

[16] M. Koehn, Solutions of the Klein-Gordon equation in an infinite square-well potential with a moving wall, EPL **100**, no.6, 60008 (2012) [arXiv:1301.0436 [math-ph]].

[17] M. Baggioli and M. Landry, Effective Field Theory for Quasicrystals and Phasons Dynamics, SciPost Phys. **9**, no.5, 062 (2020) [arXiv:2008.05339 [hep-th]].

[18] T. C. Lubensky, Sriram Ramaswamy, and John Toner, Hydrodynamics of icosahedral quasicrystals. Phys. Rev. B 32, 7444, (1985).

[19] Baake, M., Grimm, U. Aperiodic Order, Cambridge University Press, (2013).

[20] Elser, V. and Sloane, N. J. A., A highly symmetric four-dimensional quasicrystal, J. Phys. A 20, 6161, (1987).

[21] Conway, J. H. and Sloane, N. J. A. Sphere Packins Lorricer and Groups (Berlin: Springer) (1988).

[22] Sadun, L.: Tilings, tiling spaces and topology, Philosophical Magazine 86, 875-881, (2006).

[23] Connes, A.: Non-commutative geometry, Boston, MA: Academic Press, (1994).

[24] Z. Li and L. Boyle, "The Penrose Tiling is a Quantum Error-Correcting Code," [arXiv:2311.13040 [quant-ph]].

[25] Amaral, M.; Chester, D.; Fang, F.; Irwin, K. Exploiting Anyonic Behavior of Quasicrystals for Topological Quantum Computing. Symmetry 2022, 14, 1780. https://doi.org/10.3390/sym14091780

[26] Barbour, J. Class. Quantum Grav. 20, 1543. (2003) https://doi.org/10.1088/0264-9381/20/8/310

[27] M. Fabbrichesi, E. Gabrielli, and G. Lanfranchi, "The Physics of the Dark Photon: A Primer," SpringerBriefs in Physics, Springer, Cham (2020), doi:10.1007/978-3-030-62519-1.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19690-y

Towards optimal score level fusion for adaptive multi-biometric authentication system

Kavita[1] · **Rajesh Rohilla**[1] · **Gurjit Singh Walia**[2]

Received: 24 January 2024 / Revised: 19 May 2024 / Accepted: 11 June 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
Multi-biometric authentication systems have been extensively investigated due to their merit over unimodal systems. Multi-biometric systems not only overcome various security and privacy concerns of unimodal systems but also enhance overall performance & reliability. However, optimal fusion of multi-biometrics to overcome environmental challenges remains an open-ended research problem that was limitedly addressed. To address this issue, optimal score level fusion is proposed for a multi-biometric authentication system which includes biometrics namely fingerprint, finger-vein, and iris. The proposed optimal score level fusion technique exploits the Whale Optimization Algorithm (WOA) to efficiently weight belief assignment of finger-vein, iris, and fingerprint classifiers. The key features of WOA namely encircling, exploration, and exploitation are exploited to obtain optimal solutions for the fusion of multi-biometrics. For this, belief assignment for the score obtained from three modalities is computed using the Denoeux model. Furthermore, belief assignments are fused using the combination rule of PCR-6 (Proportional Conflict Redistribution) to achieve adaptiveness of the system to environmental effects. The proposed multi-biometric system is evaluated over a benchmark dataset namely, VERA finger vein, FVC 2002 and FVC 2006 fingerprint, and IITD PolyU iris. The experimental evaluation shows that the proposed technique attains, on an average, accuracy of 99.33% & equal error rate (EER) of 0.27% and these results are compared with the state-of-the-art system.

Keywords Multi-biometric · Whale Optimization Algorithm (WOA) · Authentication · Optimal Fusion · Belief mass

✉ Rajesh Rohilla
rajesh@dce.ac.in

Kavita
kavbit.207@gmail.com

Gurjit Singh Walia
gurjit.walia@gmail.com

[1] Electronics & Communication, Delhi Technological University, Delhi, India

[2] SAG, DRDO, Delhi, India

Published online: 15 July 2024

1 Introduction

Biometric technologies are widely used in identifying and authenticating users. Biometric systems are very crucial for security requirements and access control. These systems utilized physical (face, iris, finger vein, fingerprint, etc.) or behavioral (voice, gait, signature, etc.) traits to recognize the authenticity of a person for verification or identification. Biometric traits cannot be used, stolen, or forged easily. Hence, biometric systems are considered more suitable as compared to the traditional security system. However, using a single biometric trait to establish the identity of a person suffers from many limitations, including noise, non-universality, intra-class variation, and spoof attacks. These limitations can be circumvented through the fusion of multiple biometric traits.

In multi-biometric systems, two or more physical or behavioral traits are combined to increase security, reliability, and resistance to spoof attacks. These integrated biometric systems provide an optimal value of equal error rate (EER) and hence improve system accuracy [1, 2]. In addition, researchers have developed many fusion techniques that can be used before or after the matching module to enhance the effectiveness of multi-biometric systems. These fusion approaches are difficult to apply due to variations in acquired biometric data and the size of derived feature vectors. Unless all modalities are related to the same kind of feature vectors that have been adjusted to a uniform proportion. Following the matching module, the fusion methods used are decision, rank, and score-level fusion. These techniques are commonly utilized in the field of multi-biometrics since they are independent of the characteristics of captured images or extracted features. Fusion is accomplished in all of these schemes only after individual modalities have been scored. Ranks are formed from the matching score in the rank-level fusion method. In decreasing order of confidence, these ranks represent a probable collection of matched identities. In the decision level fusion, only a decision from each matcher is required which is combined to produce the final decision. Due to this, a very limited amount of information is present during the fusion, which may result in false recognition. Score level fusion is used to produce similarity or dissimilarity scores for each biometric modality by comparing the traits of the probing subject to the traits of all enlisted subjects. Hence, score level fusion provides more information as compared to rank and decision level fusion to identify & authenticate the user [3, 4]. Multimodal biometric systems can be categorized into two distinct groups depending on the methods employed to collect various traits: (i) multiple traits collected using a single sensor and (ii) multiple traits collected using multiple sensors [5].

The constraints of unimodal biometric systems are effectively addressed by multimodal biometric systems. However, they can cause inconvenience for users during the enrolment and modalities acquisition process. Hence, from the perspective of the user, biometric features that are close to each other are chosen. The selection of biometric features frequently requires achieving a balance between system security and user comfort. Notably, modalities are integrated at different levels specifically at the feature level, score level, decision level, and rank level. Despite fusing several modalities at the score level is the easiest option to execute but achieving optimal fusion at the score level is still a challenge [6].

The proposed fusion model is robust against the spoofing attack as it is not possible to spoof three modalities at the same instance. Also, the fusion of multiple biometric traits is used to check the authenticity of the user in a multi-biometric system which makes it robust against spoof attacks. The proposed method is designed to provide expert discernment in distinguishing genuine subjects from impostors using an optimal score-level fusion approach. The fundamental fusion model acts as an intelligent agent with the utmost

decisiveness in identifying real users from imposters. The proposed score level-based fusion technique emphasizes optimizing the fusion of data derived from three complementing attributes: fingerprint, iris, and finger-vein. Furthermore, to make a final determination about user authentication, the proposed fusion process carefully considers the compatibility of the optimized source classifier. Below are the summarized key points of the proposed work:

- A novel adaptive approach for multi-biometric system using fingerprint, iris, and finger-vein
- Whale Optimization Algorithm (WOA) is exploited to cater to dynamic environmental challenges
- Conflicting beliefs among the classifiers are resolved using a score-level fusion model
- Outcome of the proposed system reveals low EER and high accuracy
- Proposed multi-biometric system is robust to security & spoofing attack

The paper comprises the following sections: An overview of related work and research gap has been highlighted in Section 2. Proposed fusion model core architecture, optimal score level fusion approach, classifier score computation, and performance optimization are discussed in Section 3. In Section 4, Implementation details and databases used for experimental simulations have been discussed. Also, the proposed system qualitative and quantitative performance is compared with the state-of-the-art techniques. Finally, Section 5 summarized our research findings and outlines future directions.

2 Related work and Research gap

To address the constraints associated with unimodal biometric systems, extensive research is currently focused on multi-biometric systems. It is believed that integrating similarity scores from numerous modalities is a pragmatic approach for combining different attributes. Details of recent research work for score-level fusion are summarized in Table 1, which depicts the comparison of score-level fusion across various modalities and a detailed analysis of experimental outcomes.

A multiple-instance score fusion strategy was suggested for the recognition of finger-knuckle prints [7]. In this work, the left index/middle finger and right index/middle finger of a person were combined at the score level. To generate fixed-length feature vectors from Finger-Knuckle Print (FKP) images, the combination of local binary pattern (LBP) encoding, Gabor filter convolution, and Bio-Hashing method was considered. The Bio-hashing approach provides separation between genuine and imposter. Likewise, the exploration of a score-based fusion for hand multi-biometrics was conducted in the study by Peng et al. [8]. Their research combined scores from finger shape, knuckle print, finger vein, and fingerprint attributes with binary operations. Frank, Yager, Hamanger, and Einstein product T-norms were used to normalize and combine the fused scores. The achievement of universality in biometric systems was not realized through these methods as they were constrained to a specific body part. Moreover, classifier performance optimization is not used in these fusion methodologies.

In 2014, Sim et al. [9] proposed a weight-based score-level fusion algorithm considering the non-ideal biometrics scenarios. The authors used a multimodal benchmark dataset (UTMIFM) for face and iris, UBIRIS version 2.0 for iris, and ORL for face. Additionally,

Table 1 Comprehensive overview of Multibiometric Authentication Systems based on score level fusion

References	Biometric Features	Database	Fusion Techniques	Techniques used	Remarks
[7]	Finger-knuckle- prints	Poly U-FKP	Multiple instance Score level	Bio-hashing approach	The performance Poly U-FKP dataset is analyzed.
[8]	Finger knuckle print, finger shape, finger-vein, finger-print	PolyU finger-knuckle-print, Hong Kong Polytechnic University finger image database Version 1.0, FVC2002	Score level	T-norms were used to fuse normalized similarity scores	The decidability index (6.94) value is low.
[9]	Iris and Face	UTMIFM (face & iris), UBIRISV2.0, ORL Face,	Weighted score level	Weighted sum is calculated using wavelet coefficients	The value of the decidability index (2.778) is low.
[12]	Voice and face	XM2VTS (voice), BANCA (face)	Weighted Score Level	Particle swarm optimization is used	EER is 1.06%
[13]	Face,fingerprint, voice	XM2VTS, NIST BSSR1	Score level	Class is determined by posterior probability	FAR value of 1% is considered for evaluation.
[14]	Iris, finger-vein, and finger-print	VERA finger vein, IITD PolyU iris, CASIA-Finger-print V5, and FVC 2006 fingerprint	Score level	Backtracking search optimization algorithm used for optimization	Accuracy of 98.43% and EER of 1.57%
[15]	Face, Iris, and Ear	ORL face, MMU iris, AMI ear, CASIA iris v1, IIT Delhi Iris & Ear	Score level fusion	Cuckoo search optimization, and PCR-6 combination rules	Accuracy of 98.31% and EER of 2.32%
[19]	Iris and Palmprint	CASIA Iris interval v4, CASIA Palmprint v1	Weighted Score Level	A weighted sum score level fusion technique was used	Achieved a recognition accuracy of 92.23%
[20]	Palmprint and Face	Face 94, 95, and 96, FERET, FRGC and palmprint IITD	Score level fusion	T-norms-based score fusion	Achieved 0.53% accuracy at FAR of 0.1%
[21]	Face and fingerprint	NIST BSSR1	Score level fusion	To solve that optimization problem, a genetic algorithm (GA) based method used	Achieved accuracy of 99.42%
[22]	Major and minor dorsal finger knuckle	Finger Knuckle Images Database (Version 1.0)	Score level fusion	Adaptive single-scale retinex (ASSR) approach is used	EER less than 0.60%

Table 1 (continued)

References	Biometric Features	Database	Fusion Techniques	Techniques used	Remarks
[23]	Latent fingerprint & iris	Fingerprint (SOCOFing) & CASIA-IrisV3	Score level fusion	Threshold and model-based techniques are discussed	The accuracy rate achieved after the fusion was 91.15% and 86.63%
[24]	Face, fingerprint, & signature	YALE Face, FVC2002 fingerprint, & KVKR Signature	Score and decision-level fusion	Scores are fused using the sum rule approach and normalized using z-normalization	Achieved a high recognition rate of about 99.04%
[26]	Fingerprint, Face, and Signature	Bio-secure DS2(signature, fingerprint & face,)	Score level	Image quality is used as a classifier score reliability factor	Achieve high performance (Approximate 1% EER)
[25]	Fingerprint and iris	SDUMLA-HMT	Transformation-based score level	Multimodal biometric watermarking, Lifting Wavelet Transform, and transformation-based score level fusion are used	Achieve high accuracy and robustness against various attacks.

Shekhar et al. [10] explored a joint sparse representation approach within the context of a multi-biometric system. In this investigation, each modality was assigned a weight based on the quality assessment of the acquired data, and non-linearity was addressed through kernels. Despite incorporating the optimized classifier scores with quality measures, the optimal fusion of iris and fingerprint modalities was addressed partially.

Mukherjee et al. [11] proposed a multimodal biometric system based on differential evolution score-level fusion approach. In which adaptive weight and exponent-based function mapping combine matching scores from different biometric sources into a single score for classifier decision-making. Differential evolution algorithm was used to adjust tunable parameters and to minimize the overlap between the frequency distributions of genuine and imposter scores. The gaussian kernel density method is used to estimate the frequency distributions for greater accuracy.

Mezai and Hachouf [12] introduced a new technique instead of directly integrating classifier scores. They transformed individual classifier scores into belief masses by integrating measures from both facial and voice features. Particle Swarm Optimization (PSO) was used to compute confidence factors for individual classifiers. The Dempster-Shafer theory and proportional conflict redistribution (PCR-5) combination rules were used to determine user classes.

In 2016, Liang et al. [13] introduced a new technique in the context of a multimodal biometric system. In this, a probability-based score fusion algorithm known as an order-preserving tree, which is non-parametric, was presented. During the normalization stage, classifier scores were converted to posterior probabilities. These probabilities were merged to generate the final probability at the merge stage. In a dynamic environment, the effectiveness of most multimodal biometric systems is often compromised due to conflicting classifier scores, which in turn diminishes the system robustness and reliability. Addressing this challenge, Walia et al. [14] introduced a multi-biometric system that relies on an optimal score-level fusion approach. In their research, they achieved optimal integration of three complementary biometric traits and to enhance the performance of individual classifiers Backtracking Search Optimization Algorithm (BSA) was used. Further, any conflicting beliefs originating from these individual classifiers are efficiently resolved through the use of PCR-6 rules to improve template security and robustness in biometric systems.

Further, the authors have presented a multi-biometric system that combines multiple features such as face, iris, and ear [15]. The authors presented important components for creating cancellable biometric characteristics, and they used the cuckoo search optimization method to produce optimal belief masses for classifiers. Finally, the fusion of classifier beliefs is carried out using proportional conflict redistribution (PCR-6) rules. In 2019, Gupta et al. [16] presented a score level fusion approach to resolve the issue of score conflict between individual classifiers. They utilized a grasshopper optimization algorithm to optimize the score of individual classifiers and scores are fused using proportional conflict redistribution (PCR-6) rules to obtain a concurrent solution. The system shows high performance, reliability, and robustness against the dynamic environment. However, it utilized only two biometric modalities.

In 2020, Abderrahmane et al. [17] proposed score level fusion approach based on the weighted quasi-arithmetic mean (WQAM). The proposed approach contains the properties of both quasi-arithmetic mean and weighted mean. It was calculated using various trigonometric functions and there is no need for a learning process. The WQAM fusion technique outperforms earlier proposed score fusion algorithms that used transformation, classification, and density estimation approaches. In [18] Gupta et al. proposed an optimal score fusion model for multiple biometric sources viz. finger, face, and iris. This system

intelligently combines scores from individual classifiers to address the challenges associated with dynamic environments, spoofing attacks, noisy input images, and other limitations. In [19] Ramachandran and Sankar suggested a score-level fusion approach for multi-biometric biometric recognition using two biometric traits namely: iris and palmprint. For feature extraction, the authors used Log-Gabor transform, Local Binary Pattern, and Histogram of Oriented Gradients. Further, scores were generated using, Chi-Square distance, Euclidean distance, and Hamming distance measures, which were combined using a weighted sum score level fusion technique. They achieved a recognition accuracy of 92.23%.Further, Rane & Bhadade [20] have presented a multi-biometric system in which they utilized palmprint and face as two biometrics features. A t-norms-based score fusion method was used, which is a binary operation and provides the fusion result in the interval 0–1.

In 2021, Ahmad et al.[21] considered score-level fusion of multimodal biometrics as an optimization problem. To solve that optimization problem, the authors used a genetic algorithm (GA) based method. The genetic algorithm-based approach utilizes a weighted Spearman foot rule distance metric to determine the distance between two score lists. The proposed method outperforms as compared to the state-of-the-art score level and rank-level fusion methods and achieves an accuracy of 99.42%.In addition, Hammouche et al. [22] presented a novel and effective method for authentication, which involves score-level fusion of Minor and major dorsal finger knuckle patterns. They extracted reflectance and illumination information from these traits using the Adaptive Single-Scale Retinex (ASSR) technique. To extract normalized histogram features and facilitate the matching stage, the authors employed a binarized statistical image features (BSSF) approach along with the Cosine Mahalanobis distance.

A multi-biometric identification system has been presented by Shreya and Chatterjee [23] in which threshold and model-based techniques were discussed. The system used two biometric traits namely latent fingerprint and iris, which are fused using the score-level fusion approach. In the threshold-based approach identification rate achieved was 80%, while in the model-based approach, it was 87.3% which was better than the threshold-based approach. The accuracy rate achieved after the fusion was 91.15% and 86.63% for the sum and product rule respectively. Kazi et al. [24] have proposed a multi-biometric system in which biometric traits face, fingerprint, and signature were fused. The authors used similarity and dissimilarity measures to calculate the score, which are fused using the sum rule approach. Normalization is performed using z- normalization method. A high recognition rate of about 99.04% using the score-level fusion approach is obtained.

To prevent online image theft Fernandez & Pandian [25] have proposed a system in which iris and fingerprint features are used as the watermark. They used registration phase to generate the watermark from the features of iris and fingerprint, and verification phase was used when an ownership dispute emerged on the image. The authors used multimodal biometric watermarking, Lifting Wavelet Transform, and transformation-based score level fusion using triangular norms to achieve better accuracy and robustness against the various attacks. Furthermore, an exploration into enhancing the effectiveness of a score-based multi-biometric fusion model through the utilization of a quality index was conducted [26]. This approach involved evaluating the reliability of input sensors based on the quality of the input image. In this study, the authors introduced the concept of an uncertainty factor to quantify the level of reliability exhibited by the source classifiers.

In sum, the literature review shows that most of the research work is performed on genetic and swarm-based algorithms which are iteration-based and used to solve complex problems but genetic algorithms are relatively expensive and have a slow convergence

rate. Score-level fusion is used by most researchers and ideal integration of more than two modalities was mostly studied in a limited way. To resolve these issues, the proposed work utilized a whale optimization algorithm (WOA) which is a population-based meta-heuristic algorithm that solves high-dimensional problems. It is easy to use and known for its simplicity and fast convergence. WOA is a global optimizer as it includes exploration and exploitation to search the optimal positions. In addition, most biometric-based authentication work is limited to one or two biometric traits which leads to security concerns and spoofing attacks. This study aims to develop an appropriate fusion model for high-security applications.

The fundamental framework of the proposed approach is detailed in the next section. Feature extraction method and proposed score-level fusion approach are also discussed in detail.

3 Core design of proposed multi-biometric system

The proposed method utilizes a hybrid approach that uses a highly competitive meta-heuristic algorithm-Whale optimization algorithm (WOA) along with proportional conflict redistribution (PCR-6) rules for optimal score fusion. Figure 1 illustrates the comprehensive architecture of the proposed multi-biometric system. The proposed score-based fusion model comprises four stages: conversion of match score into belief mass, confidence factor estimation using WOA, optimal combination of belief mass using Dezert-Smarandache Theory (DSmT) based PCR-6 rule, and decision model.

As shown in Fig. 1, three biometric traits namely iris, fingerprint, and finger-vein are taken as input and corresponding feature vectors are extracted. Comparison of the extracted feature vectors is performed with their templates stored in the database to obtain match scores S_i, S_{fv}, and S_{fp} for iris, finger-vein, and fingerprint respectively. These scores are further processed using WOA and DsmT-based PCR-6 rules for optimal combination. Whale Optimization Algorithm (WOA) is used to calculate optimized confidence values

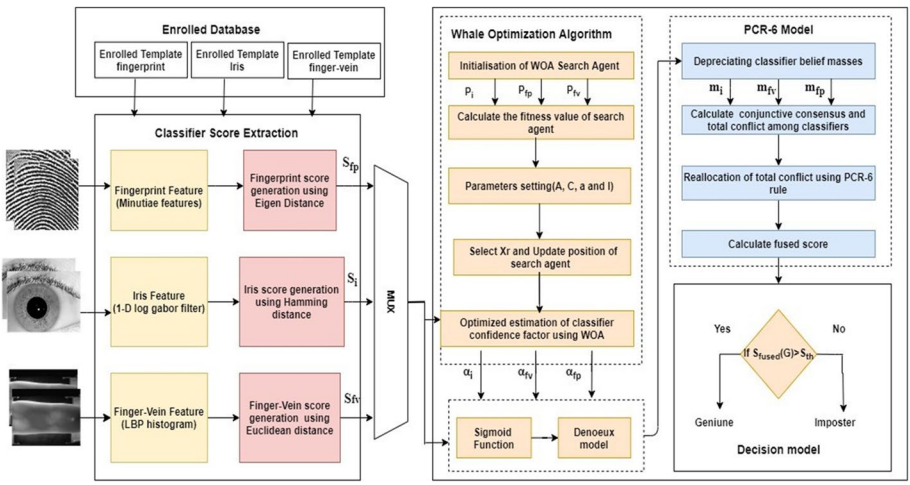


Fig. 1 Proposed Multi-biometric system using Whale Optimization Algorithm and PCR-6 based fusion rules

for individual classifiers [27]. These confidence values act as a confidence factor for the estimated match scores. The match scores are converted into the belief masses utilizing the Denoeux model. These belief masses are subjected to the DSmT-based PCR-6 rule model to resolve conflicts among individual classifier beliefs [28, 29]. Total conflict and conjunctive consensus are also determined for each classifier. The total conflict consisting of six partial conflicts is redistributed among the three classifiers and final belief mass is calculated for the entire system. Lastly, in the decision model, the final belief mass is compared with a predefined threshold value.

The proposed system demonstrates optimal performance across diverse scenarios. For example, in the case of concurrent classifiers, classifiers are boosted or suppressed. In the case of discordant classifiers, our proposed fusion model uses state-of-the-art conflict resolution. The following subsection presents the details of the proposed optimal fusion model along with the feature extraction methods.

3.1 Multi-biometric feature extraction

The three biometric features viz. fingerprint, finger-vein, and iris are selected based on their suitability and convenience of acquisition. Approaches used for the feature extraction of these modalities are as follows:

a) *Iris Feature Extraction*

Iris features are extracted using the approach by Monwar & Gavrilova [30]. Due to a precise localization of the iris region, this method has high accuracy. In this, the acquired iris image is initially processed to localize the region of interest, and after that rubber sheet model [31] is applied to the region of interest. Further, the obtained iris image undergoes convolution with 1-D log Gabor wavelets. The accuracy is ensured by filtering out the noise from the convolved pattern and the filtered pattern is transformed into phase-quantized bits. Further, the match score of the iris classifier S_i is computed by calculating the Hamming distance between the stored template and acquired phase-quantized bits using Eq. (1). The symbol $\odot$ used in Eq. (1) represents the convolved phase quantized bits of the iris.

$$S_i = \frac{\left| T_{ia} \odot T_{is} - HD_{s,a} \right|}{T_{ia} \odot T_{is}} \tag{1}$$

wherein, $HD_{s,a}$ indicate Hamming distance and is calculated using Eq. (2). T_{ia} and T_{is} indicate acquired and stored iris templates, respectively.

$$HD_{s,a} = T_{is} \oplus T_{ia} \tag{2}$$

b) *Finger-vein feature extraction*

Finger-vein features are extracted using the technique described by Peng et al. [8]. This approach achieves high accuracy because of the efficient filtering vein pattern. For this, the acquired image of a finger vein is normalized, and after that Gabor wavelet kernels are used to filter and create a map, which is known as the Gabor binary pattern map. Further, this map is divided into mutually exclusive rectangular blocks to create separate local binary pattern (LBP) histograms which are concatenated to obtain the final feature set. Further,

using Eq. (3) LBP histogram of the acquired vein image is compared against the stored templates to obtain the match score of the finger-vein classifier.

$$S_{fv} = \sum_{r} \frac{\sum_{u} \sum_{v} \min(h^{s}_{x,y,z}, h^{a}_{x,y,z})}{\sum_{u} \sum_{v} h^{s}_{x,y,z}} \tag{3}$$

where $h^{s}_{x,y,z}$ and $h^{a}_{x,y,z}$ represents the local binary pattern histogram of the stored feature set and the acquired image.

c) *Fingerprint feature extraction*

The feature extraction for fingerprint modality is attained via the minutiae extraction method presented by Ravi et al. [32]. The fuzzy rules-based approach [33] is used to filter out the false minutiae by estimating the threshold distance between terminations and the angle deviation across synchronized directions.

Further, to perform matching of legitimate minutiae pairs, ridge ending, and bifurcations are matched. The orientation of the fingerprint minutiae is considered throughout the matching process of the extracted characteristics The match score determines the similarity of two fingerprint features using Eq. (4).

$$S_{fp} = \frac{N^{2}_{match}}{N_{s} N_{a}} \tag{4}$$

where,

N_s represents the count of minutiae points in the stored template;
N_a denotes the count of matched minutiae points in the acquired template;
N_{match} represents the total number of matched minutiae pairs;

To achieve an optimal decision, the computed matched score or likelihood value of these classifiers, S_i, S_{fv}, and S_{fp} are fed into the proposed score level fusion model. The fundamental structure of the proposed model is elaborated in the next part.

3.2 Proposed score level fusion model

In a biometric system, multi-biometric fusion is essential for optimal decision-making. To integrate scores received from three biometric traits, this study provides a four-stage model of optimal fusion. In addition to performing concurrent classifier boosting and suppression, the proposed fusion model also successfully resolves conflicts between the discordant classifiers. The next subsection provides specifics of the suggested fusion model.

3.2.1 Conversion of match score into belief mass

Authentication step in a biometric system comprises a two-class problem. These classes are based on genuine and imposter users, here $\theta_{genuine}$, $\theta_{imposter}$ denoted as respectively. Therefore, based on Shafer's model the fusion problem is represented with the frame of discernment $\Theta = \{ \theta_{genuine}, \theta_{imposter} \}$. In step 1, the match scores S_y (y *i, fv, fp*) of each classifier are

transformed into belief masses m using the Denoeux model [34]. The computation of belief masses is given by

$$m_y(\theta_{genuine}) = \alpha_y * \psi(S_y) \tag{5}$$

$$m_y\left(\theta_{imposter}\right) = 1 - \alpha_y * \psi(S_y) \tag{6}$$

where $m_y(\theta_{genuine})$ and $m_y(\theta_{imposter})$ represent genuine and imposter class belief mass. α_y is the confidence factor, y$\in$iris, finger-vein, fingerprint, and ψ represent a sigmoid function or increasing function to map the scores in the range [0, 1].

The computation of the confidence factor (α_y) is performed by WOA in the next stage.

3.2.2 Confidence factor estimation using WOA

Whale Optimization Algorithm (WOA) is used for obtaining optimum values of confidence factors α_y (y$\in$*i, fv, fp*) for each classifier. WOA was proposed by Mirjalili et al. [27] which mimics the social behavior of humpback whales and is inspired by the bubble-net hunting strategy. The efficiency of WOA is evaluated by solving six structural and 29 mathematical optimization problems and, it outperforms compared to the other state-of-the-art methods. It is a nature-inspired meta-heuristic algorithm that utilizes randomly generated search agents or populations. The search for prey i.e. optimum value by search agent is divided into three phases: Encircling prey, Exploration, and Exploitation. These three phases are used for updating and improving the search agent position until the required optimum solution is not reached.

During the initialization phase, the position of the search agent is defined as P_{sa} (sa$\in$*i, fv, fp*) for each classifier. The position of each search agent is updated according to the current best position. Initially, the current best position of the search agent is defined. This newly defined best position helps other search agents to update their position towards the optimum solution. The updation of position is represented by Eq. (7) and (8)

$$\vec{D} = |\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)| \tag{7}$$

$$\vec{X}(t+1) = \overrightarrow{X^*}(t) - \vec{A} \cdot \vec{D} \tag{8}$$

where, '$\vec{A}$' and '$\vec{C}$' denotes coefficient vectors, while 't' signifies the current iteration.

Furthermore, by allocating random values to A within the interval [-1, 1], the new position of a search agent is established between the initial position of the search agent and the position of the present best search agent. '.$\vec{X}^*$' stands for the new best position vector of the best solution acquired up to the present iteration, which is continually updated if a superior solution is found. ' X ' signifies the old best position vector, | | represents the absolute value, and '·' denotes element-wise multiplication. The vectors 'A' and 'C' are computed using Eqs. (9) and (10).

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \tag{9}$$

$$\vec{C} = 2 \cdot \vec{r} \tag{10}$$

In this context, "a" progressively decreases from 2 to 0 throughout the iterations, while "r" represents a random vector within the range of [0,1].

Using the Spiral updating position technique, the distance between the search agent at (X, Y) and the prey/object at (X^*, $Y.^*$) is calculated and the position of the search agent in each iteration is updated using Eq. (11)

$$\vec{X}(t+1) = \overrightarrow{D'} \cdot e^{bl} \cdot \cos(2\pi l) + \overrightarrow{X^*}(t) \tag{11}$$

In this equation, $\overrightarrow{D'}$ signifies the distance, calculated as $|\vec{X}^*(t) - \vec{X}(t)|$, representing the separation between the old position of the search agent and the current best position or optimum solution. The parameter "b" remains constant while "l" takes on a random value within the interval of [-1, 1]. The symbol "·" denotes element-wise multiplication. Further in each iteration, the search agent element with the lowest fitness value is determined for each classifier. For this, Eq. (12) calculates the fitness value F_z to verify the optimality of the best position or solution obtained for each classifier. This value depends on the costs associated with false acceptance and rejection as well as the rates associated with each.

$$F_z = C_{fa} * FAR_z + C_{fr} * FRR_z \tag{12}$$

where, F_z denotes the fitness values of search agent position, C_{fa} and C_{fr} are the cost of false acceptance and false rejection respectively, FAR_Z and FRR_Z represent false acceptance and false rejection rate.

The lowest fitness values for each classifier act as the optimal values of confidence factor (α_y) which is used to convert individual classifier scores into belief masses by utilizing Eq. (5) and (6).

3.2.3 Optimal score level combination of belief mass

In the proposed work, the problem of multiple classifier fusion is described using Dempster Shafer's theory, with the frame of discernment $\Theta = \{ \theta_{genuine}, \theta_{imposter} \}$ stated over two focal elements. The belief masses for the frame of discernment are combined for multiple classifiers using DSmT-based PCR-6 rules [28, 29]. Proportional Conflict Redistribution(PCR-6) rule was developed by A. Martin and C. Osswald in 2006 as an alternative to the PCR-5 rule. PCR-6 is used when the number of sources or inputs to combine exceeds two. It is used to resolve the conflict among highly divergent evidence and comprises the estimation of conjunctive consensus, partial conflict, and total conflict. The proposed fusion model consists of three source (iris, fingerprint and finger-vein) system. In optimal combination of belief masses, the following steps are used:

a) Firstly, Conjunctive consensus among belief masses is calculated for the focal element $\theta_{genuine}$ as m_{cc} (θ_{genuine}) using Eq. (13).

$$m_{cc}(\theta_{\text{genuine}}) = \prod_{k=1}^{N} m_{s(k)}(\theta_{\text{genuine}}) \tag{13}$$

where N = 3 denotes the number of sources.

b) Secondly, the overall conflict among belief masses, m_{tc} ($\theta_{genuine} \cap \theta_{imposter}$) is calculated using Eq. (14), and it is composed of six partially conflicting masses.

$$m_{tc}\left(\theta_{\text{genuine}} \cap \theta_{\text{imposter}}\right) = \sum_{j=1}^{2} m_{s(1)}(Y_j) \sum \left\{ \begin{array}{l} c \bigcap_{n=2}^{3} Y(j) \cap Y_{S(n) \equiv \phi} \\ (Y_{(j)}, Y_{S(n) \in \Theta}) \end{array} \right. \prod_{k=2}^{3} m_{s(k)}(Y_{S(k)}) \tag{14}$$

These partially conflicting masses are redistributed using Eq. (15) in proportion to the mass of the focal elements that comprise the total conflict.

$$m_{rd}(\theta_{\text{genuine}}) = \sum_{\left\{ \begin{array}{l} ck = 1 \\ G \in \Theta \end{array} \right.}^{3} m_{S(k)}(\theta_{\text{genuine}})^2 \sum_{\left\{ \begin{array}{l} c \bigcap_{n=1}^{N-1} Y_{\sigma k(n) \cap \theta_{\text{genuine}} \equiv \phi} \\ Y_{\sigma k(1) \ldots\ldots Y_{\sigma k(N-1) \in \Theta}} \end{array} \right.} \left\{ cc \times \left[\frac{\prod_{j=1}^{2} m_{S(\sigma k(j))} \left(Y_{S(\sigma k(j))}\right)}{m_k(\theta_{\text{genuine}}) + \sum_{j=1}^{2} m_{S(\sigma(j))}(Y_{S(\sigma k(j))}} \right] \right. \tag{15}$$

where $m_k\left(\theta_{\text{genuine}}\right) + m_{S(\sigma k(j))}(Y_{S(\sigma K(j))}) \neq 0$ and the value of $\sigma_{k(j)}$ is calculated using Eq. (16)

$$\sigma_k(j) = \left\{ \begin{array}{l} j, if \;\; j < k \\ j+1, if \;\; j \geq k \end{array} \right. \tag{16}$$

c) Lastly, the overall belief mass is calculated for the genuine class using Eq. (17) by adding redistributed masses to conjunctive consensus.

$$S_{fused}(\theta_{genuine}) = m_{rd}(\theta_{genuine}) + m_{cc}(\theta_{genuine}) \tag{17}$$

3.2.4 Decision model

In this model, the ideal threshold S_{th} is compared with the final belief mass S_{fused} (G) to conclude. The estimation of the ideal threshold value is essential to achieve higher accuracy in a biometric system. Any variation from the ideal value can reduce the accuracy and durability of the biometric system. If the threshold value is lower than the optimal value, it may cause a large number of false acceptances (FAR). Choosing a higher threshold value rather than the optimal value results in a significant number of false rejection rate (FRR).

$$\text{Decision} = \left\{ \begin{array}{l} geniune; S_{fused}\left(\theta_{genuine}\right) \geq S_{th} \\ imposter; otherwise \end{array} \right. \tag{18}$$

Therefore, to reduce the risk of both excessive false acceptances and false rejections, the optimal threshold value of the system, denoted as S_{th}, is chosen to achieve an equal error rate. In the presented research, the experimentally validated optimal threshold value S_{th} is 0.53. Algorithm 1 outlines the pseudo-code for the proposed multi-biometric system. To validate this method, the experiment was conducted using benchmark datasets. The upcoming section provides comprehensive information about the datasets used and the implementation process for the proposed method.

Algorithm 1 Pseudo-code of proposed fusion approach

[Genuine, Imposter] = Optimal score fusion [$Image_i$, $Image_{fv}$, $Image_{fp}$]

1) Create feature vectors i.e. T_{ic}, $h^c{}_{x,y,z}$, N_c.
2) Utilizing Eqs. (1), (3), & (4), determine the S_i, S_{fv}, and S_{fp} similarity scores.
3) Stage I: Optimization of the classifier score
    a. Using Eq. (5), compute the belief masses [$m_i(\theta_{genuine})$, $m_{fv}(\theta_{genuine})$, $m_{fp}(\theta_{genuine})$].
    b. Initialize the position of the search agents
    c. Determine the fitness values [$F_z(i)$, $F_z(fv)$, $F_z(fp)$]
    d. Evaluate the fitness of each search agent
    e. Consider that search agent X* is the best.
    f. If current iteration< maximum number of iteration update a, A, C, l, and p.
    g. Adjust the current search agent position according to Eq. (7) for p<0.5 and |A|<1.
    h. If $|A| \geq 1$, Choose a random variable X_r and use eq (8) to update the position of the search agent.
    i. If $|p| \geq 1$, Choose a random variable X_r and use eq (11) to update the position of the search agent.
    j. Determine each search agent fitness and update the optimal position (X*) for an improved outcome
    k. Optimized estimation of the classifier confidence factor
4) Stage II:
    a. Discounting classifier belief masses
    b. Compute classifier conjunctive consensus and total conflict
    c. Redistribution of total conflict using PCR-6 rule
    d. Evaluate the fused score
5) Model of Determination:

    If ($S_{fused}(G) >= S_{th}$), Genuine

    Else Imposter

4 Database and experimental validation

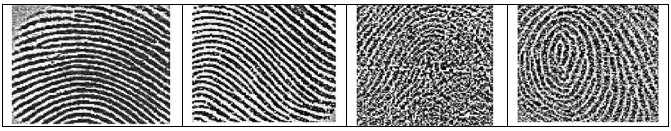
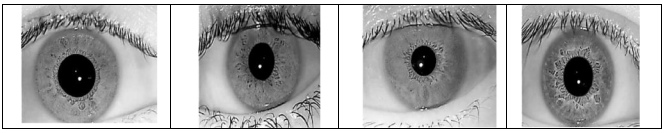
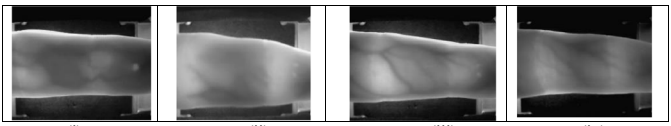
This section contains a description of the benchmark datasets, experimental evaluation, and results obtained on the proposed fusion model. Also, to check the robustness and stability, the proposed system is compared with the currently used approaches.

4.1 Database

Different benchmark datasets are employed to create multimodal datasets that are used to validate the proposed technique. Specifically, four benchmark datasets: FVC 2006 fingerprint [35], FVC 2002 fingerprint [36], IITD PolyU Iris [37], and VERA finger vein [38] are used to create two chimeric datasets. These chimeric datasets are created by a unique combination of modalities (fingerprint, finger-vein and iris) obtained from four benchmark datasets.

These datasets are often used in the evaluation of biometric systems. Table 2 presents sample images from the selected datasets. The FVC2006 fingerprint dataset consists of fingerprint images from 140 subjects, with each subject having 12 images captured at different orientations. Each image is sized at 96×96 pixels. On the other hand, the FVC 2002 fingerprint database includes multiple images of each finger from different subjects, acquired at various orientations, and these images have dimensions of 328×356 pixels. The images acquired in indoor conditions employing a digital CMOS camera are included in the IITD

Table 2 Example images from benchmark datasets for experimental evaluation

Database	Example Imag
a) Fingerprint	 (i) (ii) (iii) (iv) FVC 2006 fingerprint dataset
b) Iris	 (i) (ii) (iii) (iv) IITD PolyU Iris dataset
c) Finger vein	 (i) (ii) (iii) (iv) VERA finger-vein datas

PolyU iris database. Within this dataset, there are five images available of each subject for both left and right eyes, with each image sized at 320×240 pixels. The VERA finger-vein dataset consists of 110 users and contains 6 images for the index, ring, and middle, fingers of both the right and left hands, encompassing a mix of indoor and outdoor environments. Each image within this database is sized at 665×250 pixels. In the experimental validation section, the proposed multimodal biometric system is utilized for reference and to validate its performance with the aid of two chimeric datasets: Dataset 1 and Dataset 2. Dataset 1 is composed of samples from 50 different subjects, drawing from FVC2006 fingerprint, IITD PolyU Iris, and VERA finger-vein datasets. Initially, 50 unique subjects were allocated from the iris dataset to 50 unique subjects from the finger vein dataset. Subsequently, 50 unique subjects were assigned from the fingerprint dataset to the subjects from the previously combined dataset. This process results in the creation of a virtual multimodal dataset comprising 50 unique subjects. Similar to Database 1, Database 2 is formed by selecting 40 distinct subjects from the IITD PolyU Iris, FVC2002 fingerprint, and VERA finger vein datasets. The subjects included in Database 1 are entirely separate and do not overlap with the subjects chosen for Database 2.

4.2 Experimental validation

The effectiveness and performance of the proposed multi-biometric system are developed and simulated on the MATLAB R2018 platform. The test platform (Intel i5 processor and

16 GB RAM) took 23 s to process dataset 50 identities over 25 algorithm runs. This shows that the system takes 25.01 ms on average, to authenticate each identity. The computational time for 50 identities is 25.01 ms, which is low compared to Particle Swarm Optimization (PSO), Sum Rule, and Backtracking Search Optimization Algorithm (BSA) methods. The experimental results are achieved through the fivefold cross-validation of the proposed fusion model. For this purpose, subject samples are separated into four enrolled samples and one test sample for each subject. The process is repeated five times, to ensure each sample is utilized as the test sample exactly once.The achievement of low computational demands can be primarily attributed to the optimization of classifier scores through the WOA. A chimeric multimodal dataset, which combines VERA finger vein, FVC 2002 fingerprint, FVC 2006 fingerprint, and IITD PolyU iris datasets, is used for evaluating the performance of the multi-biometric fusion model. The evaluation process is categorized into two parts: Quantitative evaluation and Qualitative evaluation. In qualitative evaluation, three biometric modalities fingerprint, iris, and finger-vein are fused to produce a final match score, which is then compared with the individual classifier match scores. Conversely, in quantitative evaluation, two performance metrics, decidability index (DI) and equal error rate (EER) are calculated and compared with other state-of-the-art techniques.

4.2.1 Qualitative evaluation

The qualitative evaluation of the proposed multimodal system is conducted through the utilization of a chimeric dataset. The optimal combination of three classifiers involves two significant stages in the proposed muti-biometric fusion scheme. The first significant stage involves determining the lowest fitness values for each classifier which in turn act as an optimized confidence factor (α_y). These confidence factors of respective classifiers are then converted into belief masses as discussed in Section 3.2. The larger belief masses are obtained for genuine subjects while smaller belief masses are for imposter subjects of each classifier. Further, these belief masses are combined/fused to obtain the fused mass. For this, Fig. 2a and b illustrate the process of resolving conflicts between belief masses from conflicting classifiers. When larger values of belief masses are obtained from divergent classifiers, they tend to result in larger values of fused masses, particularly in the case of genuine subjects. In simpler terms, substantial belief masses from individual classifiers validate a subject authenticity and consequently contribute to an increased overall belief mass.

As depicted in Fig. 2a, consider subject 8, the final fused mass surpasses the individual classifier belief masses due to concurrent agreement among the classifiers and hence illustrates a scenario of belief mass enhancement. Conversely, when all classifiers collectively indicate that the subject is unauthentic, the final fused mass is diminished, indicating belief mass suppression. Individual classifier belief masses, as depicted in Fig. 2b, are considerably smaller than the final fused mass. In situations where classifiers provide conflicting information, Eq. (14) is applied to assess the conflict among classifier masses, which is subsequently redistributed among the individual components based on their contribution to the conflict (Fig. 3).

The two-stage integration enhances the performance of individual classifiers and efficiently combines their belief masses. Consequently, this results in a more pronounced distinction between genuine and imposter score distributions. Figure 4, shows the score distributions of the individual classifiers and the proposed approach for Dataset 1 and Dataset 2. Figure 4a and b depict the score distribution for genuine and impostor subjects

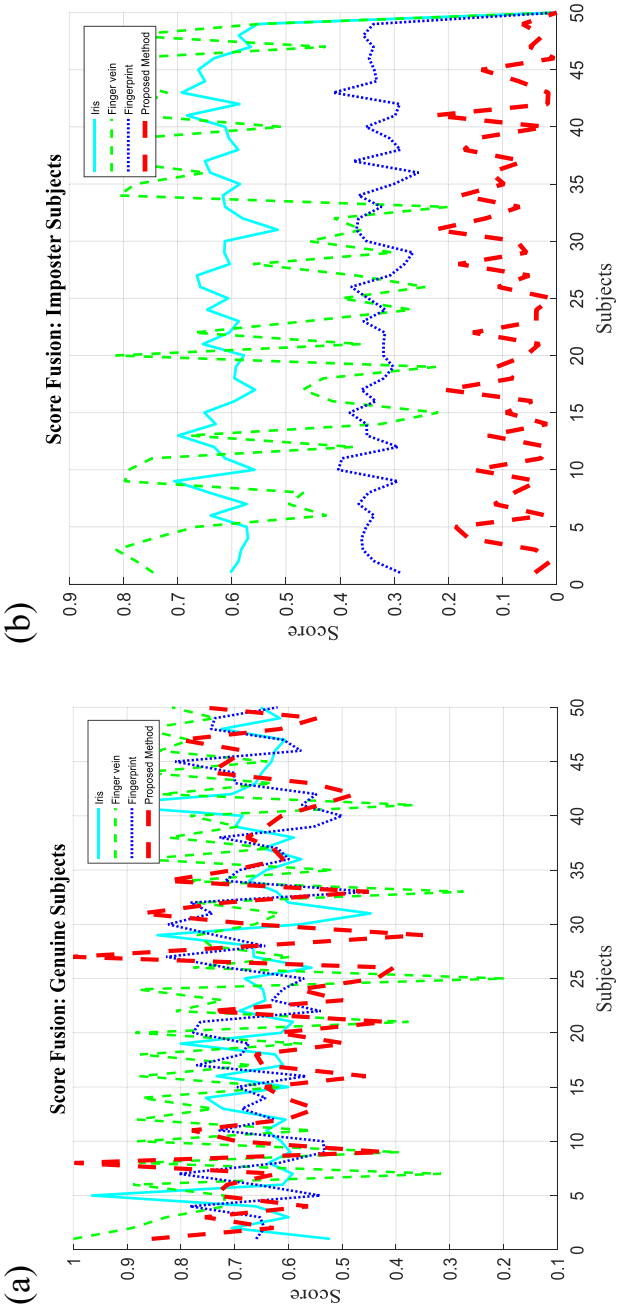


Fig. 2 Demonstration of Adaptive score fusion model vs Unimodal biometrics system for Dataset 1: (**a**) Genuine score fusion (**b**) Imposter score fusion

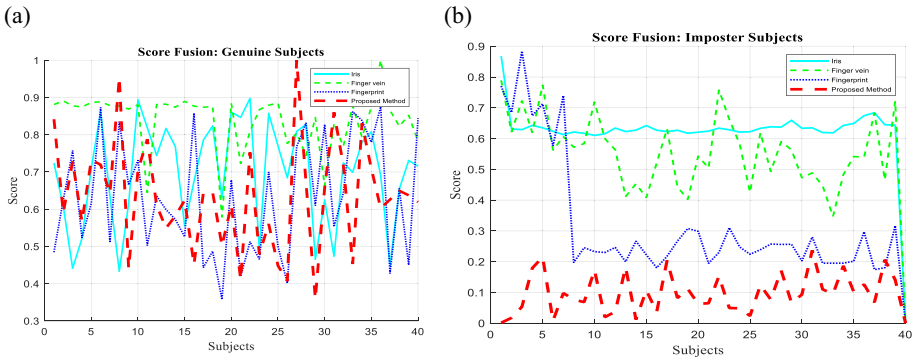


Fig. 3 Demonstration of Adaptive score fusion model vs Unimodal biometrics system for Dataset 2: (**a**) Genuine score fusion (**b**) Imposter score fusion

in the case of the fingerprint classifier, as observed in Dataset 1 and Dataset 2, respectively. Comparably, in Dataset 2, the captured fingerprint images suffer from a lack of

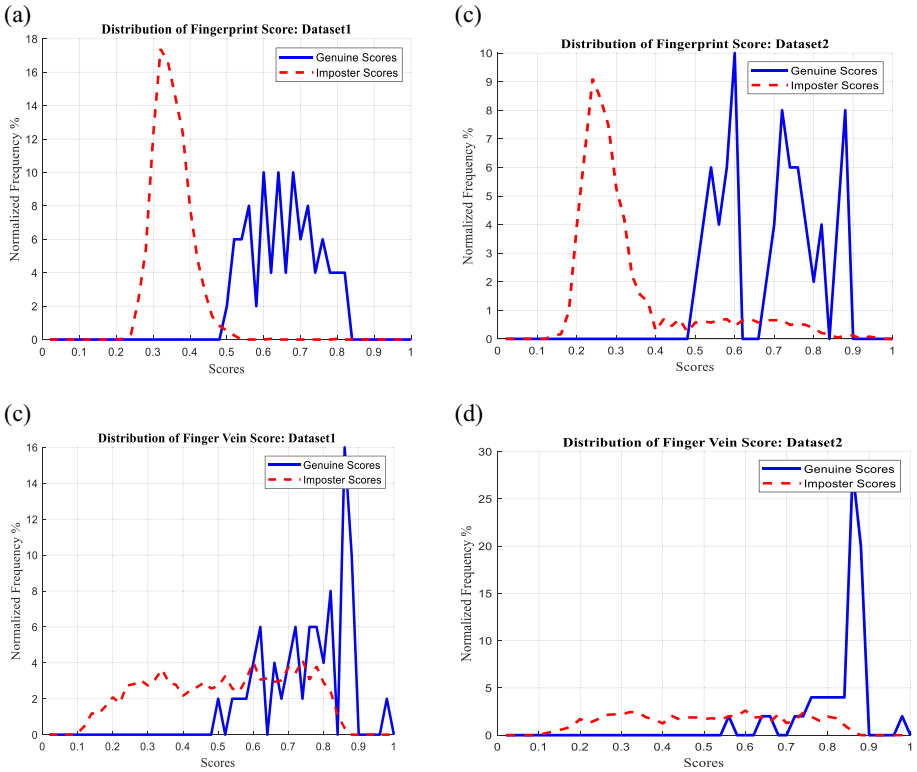


Fig. 4 Distribution of matching scores for the different classifier (**a**) fingerprint (dataset 1) and (**b**) fingerprint (dataset 2) (**c**) Finger vein (Dataset1) (**d**) Finger vein (Dataset 2) (**e**) Iris (Dataset 1) (**f**) Iris (Dataset 2) (**g**) Proposed optimal fusion model (Dataset 1) (**h**) Proposed optimal fusion model (Dataset2)

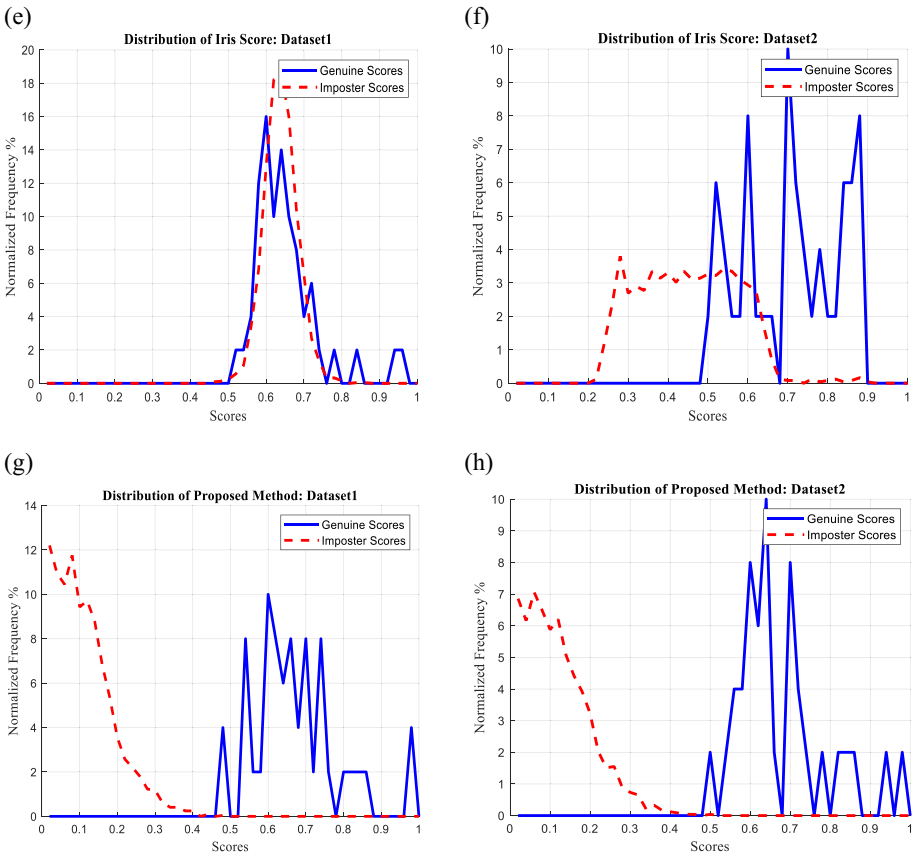


Fig. 4 (continued)

boundary information and are also affected by the surrounding environment. Consequently, this results in overlapping of score distribution for the genuine and impostor classes. Conversely, in Dataset 1, the fingerprint trait exhibits only a minor overlap between genuine and impostor scores, which can be attributed largely to the high-quality fingerprints of the subjects and the presence of distinct boundary information.

Likewise, Fig. 4c and d show the genuine and imposter score distribution for a finger-vein classifier. Even though vein patterns that are located underneath the skin are less susceptible to environmental factors, still, the distribution of imposter/ genuine scores is mainly attributed to the finger vein classifier employed for score estimation. Consequently, the false acceptance rate/ false rejection rate gets impaired and hence decreases the overall accuracy of the biometric authentication system. Genuine class and impostor class score distributions for the iris classifier for datasets 1 and dataset 2 respectively overlap with one another, as illustrated in Fig. 4e and f. Inconsistencies in iris shape, off-angle images, and blur are mostly responsible for this overlap between the score distribution of genuine and imposter classes. An increase in the false acceptance rate is caused by a small gap between the distributions of genuine and imposter class scores, which reduces the accuracy of the iris-based unimodal biometric system. However, score distributions for the genuine

Table 3 Decidability index for the individual classifier

Individual Classifiers	Decidability Index (DI)		
	Dataset 1	Dataset 2	Combined dataset (Dataset 1 and Dataset 2)
Fingerprint	4.814	6.048	5.121
Finger-vein	5.623	5.028	5.425
Iris	4.291	4.158	4.227

Table 4 Decidability index for different score-based fusion models

Score-based fusion approach	Decidability Index (DI)		
	Dataset 1	Dataset2	Combined dataset (Dataset 1 and Dataset 2)
PSO	8.901	12.632	10.420
Sum Rule	4.021	3.670	3.864
BSA	13.164	17.068	14.893
Proposed approach	4.6545	4.6351	4.643

and imposter classes of the proposed method are significantly distinct, which produces an optimal decision boundary for the two classes. Figure 4g and h, illustrate the separation of the score distribution for the proposed biometric system for dataset 1 and dataset 2, respectively. Further, the score distributions are enhanced by quantitative examination of the decidability index and equal error rate which is explained in the next section.

4.2.2 Quantitative evaluation

Biometric systems are quantitatively compared and characterized using equal error rates obtained from receiver operating characteristic (ROC) curves, as well as accuracy measures. High accuracy and low EER are desirable when aiming for the development of a high-performance biometric system. Equal error rate and decidability index are utilized to evaluate the performance of the proposed optimal fusion model. Decidability values for individual classifiers and different fusion models are shown in Tables 3 and 4. Also, the comparison of EER values of the proposed fusion model with various existing score fusion techniques namely: (i) Back-tracking Search Optimization Algorithm (BSA) [14] (ii)Particle Swarm Optimization (PSO) [39] and, (iii) Sum Rule [40] is shown in Table 5. Furthermore, the ROC curves plotted for different score fusion techniques and the proposed fusion model are shown in Fig. 5.

The proposed fusion model gives the lowest EER values, 0.3061 on dataset 1 and 0.2449 on dataset 2, as compared to the other existing fusion approaches, shown in Table 5.

Table 5 Comparison of EER values of score-based fusion models

Database	Proposed approach	BSA [14]	PSO [39]	Sum Rule [40]	Fingerprint	Finger Vein	Iris
Chimeric dataset1	0.3061	2.08	2.13	12.10	14.91	14.16	18.01
Chimeric dataset2	0.2449	0.98	1.29	4.35	10.03	13.20	16.28
Combined dataset (Dataset 1 and Dataset 2)	0.278	1.16	1.75	8.65	12.73	13.72	17.22

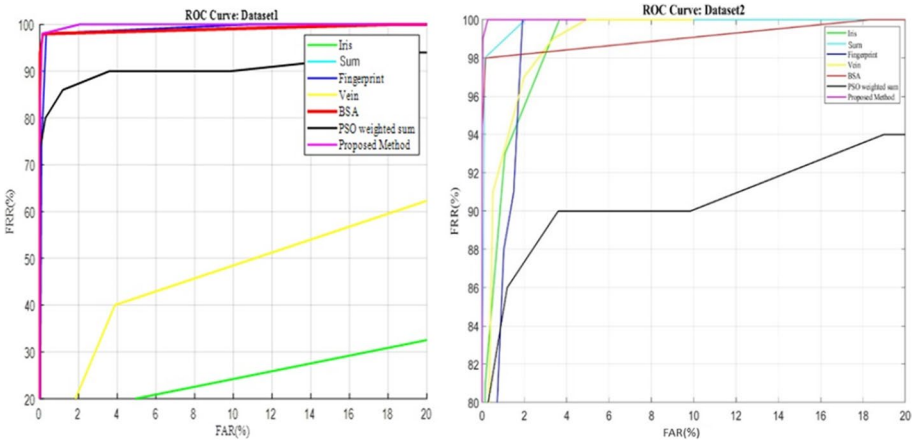


Fig. 5 Comparison of ROC curves of different fusion techniques with proposed fusion model for (**a**) Chimeric dataset1 and (**b**) Chimeric dataset2

Decidability index Decidability is a performance metric that is utilized to measure the separation between genuine and imposter score distributions. It is the difference between the means of genuine and impostor distributions relative to their variances and calculated using Eq. (19).

$$\mathrm{d} = \frac{|\mu_g - \mu_i|}{\sqrt{(\sigma_g^2 - \sigma_i^2)/2}} \tag{19}$$

where σ_i and σ_g represent standard deviations of imposter and genuine scores distribution respectively. μ_i and μ_g represent mean values of imposter and genuine scores distribution respectively.

With a high decidability value, a classifier can effectively distinguish between impostor and genuine subjects. To demonstrate, the low decidability index d of 4.814 for the fingerprint trait associated with Dataset 1 is mainly due to the inadequate subject boundary information. However, the high decidability index of d = 6.048 for dataset 2 is a result of the high quality of fingerprint images. In contrast, a uniform distribution of impostor scores leads to a low decidability index of 5.623 for Dataset 1 and 5.028 for Dataset 2 for

the finger vein classifier. Generally, the biometric systems that entrust simply on finger veins cannot distinguish between genuine and imposter subjects. Similarly, iris classifier has a low decidability index of 4.291 and 4.158 for Dataset 1 and Dataset 2, respectively, primarily attributed to the presence of off-angle images, blurriness, and inconsistencies in iris shape.

Table 4 presents a comparative analysis of the decidability index between the proposed fusion method and three other currently used fusion techniques namely: Particle Swarm Optimization (PSO), Sum Rule, and Backtracking Search Optimization Algorithm (BSA). The proposed fusion method achieved a decidability index of 4.6545 and 4.6351 for Dataset 1 and Dataset 2 respectively, which is better as compared to the Sum Rule approach.

Accuracy The accuracy measures the capacity of a system to correctly authenticate an enrolled user and is calculated using the following equation:

$$\text{Accuracy} = \frac{N_{gen}}{N_{imp} + N_{gen}} \tag{20}$$

where N_{gen} *and* N_{imp} represents the number of successful genuine and impostor attempts respectively.

Time complexity analysis Time complexity is a measure of how the execution time of an algorithm changes as the size of its input increases or the number of subjects increases. The computational efficiency of the proposed method is evaluated through the time complexity analysis of the score fusion process of different methods.

Equal error rate The Receiver Operating Characteristic (ROC) curve shown in Fig. 5, is plotted to determine the equal error rate. The curve is plotted between false rejection rate (FRR) and false acceptance rate (FAR) to compare different fusion techniques. The ROC curve for Dataset 1 in Fig. 5a demonstrates that the proposed technique requires a high FRR value for a given FAR value, resulting in high accuracy. Similarly, the ROC curve for Dataset 2 in Fig. 5b shows that the proposed technique requires a high FRR value for a given FAR value, resulting in high accuracy. The EER values for each examined fusion technique are summarized in Table 5.

The values presented for the proposed score-based fusion approach are average values acquired over 25 runs of the algorithm, which takes into account the random nature of WOA. Similarly, EER values for the PSO-weighted sum fusion approach are established as average values collected over 30 iterations of the fusion process. However, the Sum Rule fusion approach is deterministic hence, results were acquired after just one iteration of the fusion algorithm. As demonstrated in Table 5, the proposed method provides the lowest EER value for Datasets 1 and 2.

The fusion technique Sum Rule has high EER values mainly as a result of its deterministic fusion model and absence of classifier optimization. Though PSO-Weighted-Sum optimizes classifier score but conflicts between classifiers are not resolved, which leads to a high EER value. Similarly, BSA has a higher value EER as compared to the proposed approach due to the conflict between the classifiers. According to the quantitative study, individual classifiers are incapable of effectively discriminating between the genuine and imposter classes. This is mostly caused by the iris feature having an irregular shape and off-angle blur, as well as a lack of border information for finger vein and fingerprint features.

Table 6 Comparison of accuracy of different score-based fusion methods

Database	Proposed approach	Combined dataset	BSA	PSO	Sum Rule
Accuracy(%)	99.33	98.63	97.63	97.13	96.14

Table 7 Comparison of computational time for different fusion methods

Method	Proposed approach	BSA	PSO	Sum Rule
Time (milliseconds)	25.01	28.22	29.02	32.33

Therefore, to overcome these limitations the proposed optimal fusion model combines individual classifiers. Which, in turn, provides a high decidability index and low value of EER.

The key benefit of the proposed fusion model is that it can be utilized in critical applications demanding low computational time as well as ensuring varying level of security. WOA-PCR6-based fusion model performs efficiently because of the performance optimization, which is accomplished by resolving conflict among discordant classifier scores and boosting or suppressing concurrent classifiers (Tables 6 and 7).

5 Conclusion and future directions

Multi-biometric authentication systems have been extensively used due to their merit over unimodal biometric systems. Biometric modalities can be fused at feature, score, and decision level for the realization of the system however, score level fusion suffers due to the fusion of scores from different traits which are affected by dynamic environment. For this, in this manuscript, an optimal score-level fusion model has been proposed based on DSmT and WOA. The three complementary features namely fingerprint, iris, and finger-vein are considered for the realization of a multi-biometric system. The major contribution of this research work consists in evaluating the globally fit value of the confidence factor with the highly competitive nature inspired meta-heuristic algorithm WOA. Then, these factors transform the match score of three classifiers, respectively into belief mass. Individual classifier score is discounted using the Whale Optimization Algorithm and further fused using the PCR-6 Model to achieve adaptiveness to the dynamic environment. Concurrent classifier scores are booted and discordant classifier scores are suppressed leading high decidability index and hence performance optimization.

The proposed approach is evaluated on the chimeric dataset constituted from the benchmark dataset: VERA finger vein, FVC 2002, FVC 2006 fingerprint, and IITD PolyU iris of finger-vein, fingerprint, and iris respectively. The experimental result demonstrates qualitative and quantitative analysis that reveals the proposed approach surpasses other state-of-the-art fusion techniques for multimodal biometric systems. The proposed technique attains an average, accuracy of 99.33% & equal error rate (EER) of 0.27% in contrast to the current state-of-the-art system.

In the future, we will take into account the dynamic integration of uncertain information from different perspectives to provide a new paradigm for fusing scores of multiple classifiers or features of biometric traits. Also, a more efficient and effective method of integrating other factors such as the quality of biometric images, and reliability of biometric traits to improve the adaptive nature of fusion-based multimodal biometric systems.

Data availability All the details about data availability are mentioned within this manuscript. The datasets used are FVC 2006 fingerprint [28], FVC 2002 fingerprint [29], IITD PolyU Iris [30], and VERA finger vein [31].

Declarations

Ethics approval This paper does not contain any studies with animals or human participants performed by any of the authors.

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Jain AK, Ross A, Prabhakar S (2004) An Introduction to Biometric Recognition. IEEE Trans Circuits Syst Video Technol 14:4–20. https://doi.org/10.1109/TCSVT.2003.818349
2. Jain AK, Kumar A (2012) Biometric recognition: an overview. Second generation biometrics: the ethical, legal and social context, pp 49-79. https://doi.org/10.1007/978-94-007-3892-8_3
3. Unar JA, Seng WC, Abbasi A (2014) A review of biometric technology along with trends and prospects. Pattern Recognit 47:2673–2688. https://doi.org/10.1016/j.patcog.2014.01.016
4. Ross A, Jain A (2003) Information fusion in biometrics. Pattern Recognit Lett 24:2115–2125. https://doi.org/10.1016/S0167-8655(03)00079-5
5. Lumini A, Nanni L (2017) Overview of the combination of biometric matchers. Inf Fusion 33:71–85. https://doi.org/10.1016/j.inffus.2016.05.003
6. Kuncheva LI (2003) "Fuzzy" Versus "Nonfuzzy" in Combining Classifiers Designed by Boosting. IEEE Trans Fuzzy Syst 11:729–741. https://doi.org/10.1109/TFUZZ.2003.819842
7. Shariatmadar ZS, Faez K (2013) Finger-knuckle-print recognition via encoding local-binary-pattern. J Circuits, Syst Comput 22(06):1350050
8. Peng J, El-Latif AAA, Li Q, Niu X (2014) Multimodal biometric authentication based on score level fusion of finger biometrics. Optik (Stuttg) 125:6891–6897. https://doi.org/10.1016/j.ijleo.2014.07.027
9. Sim HM, Asmuni H, Hassan R, Othman RM (2014) Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images. Expert Syst Appl 41:5390–5404. https://doi.org/10.1016/j.eswa.2014.02.051
10. Shekhar S, Patel VM, Nasrabadi NM, Chellappa R (2014) Joint sparse representation for robust multimodal biometrics recognition. IEEE Trans Pattern Anal Mach Intell 36:113–126. https://doi.org/10.1109/TPAMI.2013.109
11. Mukherjee S, Pal K, Majumder BP et al (2014) Differential evolution based score level fusion for multimodal biometric systems. In: 2014 IEEE symposium on computational intelligence in biometrics and identity management (CIBIM). IEEE, pp 38–44. https://doi.org/10.1109/CIBIM.2014.7015441
12. Mezai L, Hachouf F (2015) Score-Level Fusion of Face and Voice Using Particle Swarm Optimization and Belief Functions. IEEE Trans Human-Machine Syst 45:761–772. https://doi.org/10.1109/THMS.2015.2438005
13. Liang Y, Ding X, Liu C, Xue JH (2016) Combining multiple biometric traits with an order-preserving score fusion algorithm. Neurocomputing 171:252–261. https://doi.org/10.1016/j.neucom.2015.06.039
14. Walia GS, Singh T, Singh K, Verma N (2019) Robust multimodal biometric system based on optimal score level fusion model. Expert Syst Appl 116:364–376. https://doi.org/10.1016/j.eswa.2018.08.036
15. Walia GS, Rishi S, Asthana R et al (2019) Secure multimodal biometric system based on diffused graphs and optimal score fusion. IET Biometrics 8:231–242. https://doi.org/10.1049/iet-bmt.2018.5018
16. Gupta K, Walia GS, Sharma K (2019) Multimodal biometric system using grasshopper optimization. In: 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE, pp 387–391. https://doi.org/10.1109/ICCCIS48478.2019.8974504
17. Abderrahmane H, Noubeil G, Lahcene Z et al (2020) Weighted quasi-arithmetic mean based score level fusion for multi-biometric systems. IET Biometrics 9:91–99. https://doi.org/10.1049/iet-bmt.2018.5265
18. Gupta K, Walia GS, Sharma K (2020) Quality-based adaptive score fusion approach for multimodal biometric system. Appl Intell 50:1086–1099
19. Ramachandran C, Sankar D (2020) Score level based fusion method for multimodal biometric recognition using palmprint and iris. In: 2020 Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA). IEEE, pp 281–286. https://doi.org/10.1109/ACCTHPA49271.2020.9213216

20. Rane ME, Bhadade US (2020) Multimodal score level fusion for recognition using face and palmprint. Int J Electr Eng Educ 002072092092966. https://doi.org/10.1177/0020720920929662
21. Ahmad S, Pal R, Ganivada A (2021) Score level fusion of multimodal biometrics using genetic algorithm. In: 2021 IEEE Congress on Evolutionary Computation (CEC). IEEE, pp 2242–2250. https://doi.org/10.1109/CEC45853.2021.9504927
22. Hammouche R, Attia A, Akhrouf S (2022) Score level fusion of major and minor finger knuckle patterns based symmetric sum-based rules for person authentication. Evol Syst 13:469–483. https://doi.org/10.1007/s12530-022-09430-8
23. Shreya S, Chatterjee K (2024) Latent fingerprint and Iris fusion for enhancement of performance of human identification system. Expert Syst Appl 235:121208. https://doi.org/10.1016/j.eswa.2023.121208
24. Kazi M, Kale K, Mehsen RS et al (2023) Face, fingerprint, and signature-based multimodal biometric system using score level and decision level fusion approaches. IETE J Res 1–20. https://doi.org/10.1080/03772063.2023.2217784
25. Fernandez JJPN (2023) Protection of online images against theft using robust multimodal biometric watermarking and T-norms. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-17497-x
26. Nguyen K, Denman S, Sridharan S, Fookes C (2015) Score-level multibiometric fusion based on dempster-shafer theory incorporating uncertainty factors. IEEE Trans Human-Machine Syst 45:132–140. https://doi.org/10.1109/THMS.2014.2361437
27. Mirjalili S, Lewis A (2016) The Whale Optimization Algorithm. Adv Eng Softw 95:51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008
28. Martin A, Osswald C (2006) A new generalization of the proportional conflict redistribution rule stable in terms of decision. Advances and applications of DSmT for information fusion (Collected Works): 2. Ameri- can Research Press, Rehoboth, pp 69–88
29. Smarandache F, Dezert J (2006) "Proportional conflict redistribution rules for information fusion" in Advances and Application on DSmT for Information Fusion (Collected Works): 2. American Research Press, Rehoboth, pp 3–68
30. Monwar MM, Gavrilova M (2013) Markov chain model for multimodal biometric rank fusion. Signal, Image Video Process 7:137–149. https://doi.org/10.1007/s11760-011-0226-8
31. Daugman J (2004) How iris recognition works. IEEE Trans Circuits Syst Video Technol 14(1):21–30. https://doi.org/10.1109/TCSVT.2003.818350
32. J R, Raja KB, R VK (2010) Fingerprint Recognition Using Minutia Score Matching. 1:35–42. https://doi.org/10.48550/arXiv.1001.4186
33. Surya N, Bhavanarayana K, Sathish K (2013) Removal of False Minutiae Using Fuzzy Rules 3:56–62
34. Denœux T, Masson M-H (2010) Dempster-Shafer Reasoning in Large Partially Ordered Sets: Applications in Machine Learning. pp 39–54 https://doi.org/10.1007/978-3-642-11960-6_5
35. Cappelli R, Ferrara M, Franco A, Maltoni D (2007) Fingerprint verification competition 2006. Biometric Technology Today 15(7–8):7–9
36. Maio D, Maltoni D, Cappelli R et al (2002) FVC2002: second fingerprint verification competition. In: 2002 International conference on pattern recognition. IEEE, pp 811–814. https://doi.org/10.1109/ICPR.2002.1048144
37. Kumar A, Passi A (2010) Comparison and combination of iris matchers for reliable personal authentication. Pattern Recognit 43:1016–1026. https://doi.org/10.1016/j.patcog.2009.08.016
38. Tome P, Raghavendra R, Busch C et al (2015) The 1st competition on counter measures to finger vein spoofing attacks. In: 2015 International Conference on Biometrics (ICB). IEEE, pp 513–518. https://doi.org/10.1109/ICB.2015.7139067
39. Srinivas N, Veeramachaneni K, Osadciw LA (2009) Fusing correlated data from multiple classifiers for improved biometric verification. In: 2009 12th International Conference on Information Fusion. IEEE, pp 1504–1511
40. Ghulam Mohi-ud-Din S, Bin MA, Masood H, Mumtaz M (2011) Personal identification using feature and score level fusion of palm- and fingerprints. Signal, Image Video Process 5:477–483. https://doi.org/10.1007/s11760-011-0251-7

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19801-9

UDCT: lung Cancer detection and classification using U-net and DARTS for medical CT images

Aakanksha Gupta[1] · **Ashwni Kumar**[1] · **Kamakshi Rautela**[2]

Received: 27 July 2023 / Revised: 25 June 2024 / Accepted: 27 June 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
Lung cancer is the most fatal disease in recent times. Early detection of the same is very crucial and challenging task. Therefore, proper diagnostic and treatment strategies should be employed for the automatic detection of nodules from Computed Tomography (CT) images. In this direction this paper introduces a novel UDCT, a new and effective framework for detecting and classifying lung cancer based on CT images. The proposed framework exploits Modified U-Net architecture that utilizes the traditional U-Net architecture for multi-scale feature extraction with the DARTS (Differentiable Architecture Search) which utilizes the concept of continuous relaxation of the architecture representation with efficient gradient descent search for efficient classification of lung cancer. In addition to this, the proposed framework, incorporates Multilevel Otsu thresholding for image preprocessing and to further segment the lung nodule from the CT scan images. Furthermore, the modified U-Net is capable to detect and classify lung nodule effectively and efficiently. To evaluate the performance of the proposed UDCT framework, extensive experiments are conducted on LIDC-IDRI and IQ-OTH/NCCD CT image datasets that provided state-of-the-art results. The proposed UDCT framework provided an accuracy of 95.01% on LIDC-IDRI dataset and 96.82% on IQ-OTH/NCCD dataset. An ablation study is also conducted that validated the efficiency and efficacy of the proposed UDCT framework.

Keywords U-net · DARTS · Lung cancer · CT images · Biomedical engineering · Segmentation · Thresholding

✉ Aakanksha Gupta
akanksha001phd20@igdtuw.ac.in

Ashwni Kumar
ashwanikumar@igdtuw.ac.in

Kamakshi Rautela
kamakshi_2k19phdec06@dtu.ac.in

[1] Department of ECE, Indira Gandhi Delhi Technical University for Women, Delhi, India

[2] Department of ECE, Delhi Technological University, Delhi, India

Published online: 13 July 2024

Abbreviations

CT	Computed Tomography
MRI	Magnetic resonance imaging
PET	Positron emission tomography
CXR	Chest radiography
DARTS	Differentiable Architecture Search
GPU	Graphics Processing Unit
LIDC-IDRI	Lung Image Database Consortium and Image Database Resource Initiative
AUC & ROC	Area Under Curve & Receiver Operating Characteristics

Variables

N & A	Normal & Abnormal classes
X & L	Threshold value & Grayscale level
p	One pixel value
Pq_a^o	Distributional probability, where 'o' is component value and $'q_a^o\ '$ is histogram.
$H_0^o(X)$ and $H_1^o(X)$	Distributional Probabilities of classes 'N' & 'A' respectively.
μ & σ	Sum and Variance of the classes.
$Z^{(u,v)}$	Output of concatenation approach, where $v-th$ is the convolution in $u-th$ layer module.
M(•), D(•), S(•) and [•]	Cascaded operation, Down-sampling function, Up-sampling function, and Channel concatenation operation respectively.
$b^{(g,h)}$	Collection of operations including convolution and max pooling
$\gamma^{(g,h)}$	Vector Quantity

1 Introduction

Lung cancer, a highly lethal disease, is responsible for approximately one out of every five cancer-related deaths. It is characterized by the presence of solid masses of tissue known as pulmonary tumor that develop in and around the lungs [1–3]. Lung cancer is a major cause of mortality for males, females, and transgender individuals worldwide, claiming around five million lives annually. The WHO (World Health Organization) estimates indicate that lung cancer causes millions of deaths each year. Unfortunately, more than 80% of lung cancer patients face a poor prognosis due to the advanced stage of the disease, which often renders surgery ineffective at the time of diagnosis [1, 4].

Detecting lung cancer early is crucial for effective treatment and saving lives since symptoms typically appear only when the disease has progressed significantly [5, 6]. Manual interpretation of numerous medical images used for lung cancer diagnosis can be time-consuming, labour-intensive, and prone to inaccuracies. Different methods such as X-rays, CT scans, MRI, PET, and breath analysis can be employed for lung cancer treatment [7, 8]. MRI and PET have limitations in identifying and staging lung cancer, while CXR (Chest radiography) and CT imaging techniques are prone to radiation exposure. However, CT imaging has become the preferred method for diagnosing lung cancer due to its ability to provide detailed information about the location and size of nodules [9]. Low-dose CT screening has been effective in detecting early-stage cancer tumours. In the past, image-processing

techniques have been used for diagnosing lung cancer, and machine-learning algorithms have played a role in analyzing CT scans and classifying lung nodules [9]. Feature extraction is a critical step in implementing machine learning, but existing methods often rely on manually created features, leading to lower accuracy. This limitation hinders computer-aided detection (CAD) systems from achieving optimal performance [10]. To address this issue, researchers have proposed the use of Artificial Intelligence, particularly deep learning. Deep learning algorithms can automatically learn important features from observed data during training, eliminating the need for manual feature creation and enabling end-to-end detection in CAD systems [11]. Convolutional Neural Networks (CNN) have shown superior performance compared to other deep learning networks in this context [12, 13].

Deep learning methods have shown promising results in classification and segmentation. Deep learning approaches, however, still face numerous challenges in the identification of lung cancer. First, the robustness of the prediction model is a challenging task as these models rely on large sets of annotated medical pictures that are contributed by seasoned radiologists, to meet their training requirements. Furthermore, great interpretability is necessary for the clinical application of deep learning; however, the learnt features cannot be satisfactorily explained by existing deep learning algorithms. Additionally, the existing approaches are not very applicable in clinical practice, mainly due to the inexperience of non-medical investigators in choosing more pertinent clinical outcomes. Therefore, the primary objectives of this study are to improve performance, reduce diagnostic time, and identify the location and characteristics of lung nodules. The goal is to develop a deep learning architecture specifically for early diagnosis and classification of lung cancer using deep learning techniques. In this direction, this paper presents a novel UDCT structure that exploits the concept of DARTS and U-Net with multi-level Otsu thresholding for easy detection of lung nodules. A basic block diagram representation is depicted in Fig. 1. Further, the proposed UDCT can effectively classify lung cancer into benign or malignant nodules.

The main Contributions of this work are:

i. This paper presents a lung cancer detection and classification framework namely, UDCT that efficiently detects and classifies lung cancer from CT images.
ii. The proposed framework exploits Multi-level Otsu thresholding on pre-processed images to efficiently extract features from CT images. Further, the extracted features are utilized by the U-Net architecture for the segmentation of CT images.
iii. Further, the proposed U-Net architecture is fused with DARTS (Differentiable Architecture Search) that uses a simplified gradient descent search in conjunction with the concept of continuous relaxation to describe architecture for efficient classification of lung cancer.
iv. To confirm the efficacy of the suggested UDCT framework, extensive experiments are conducted on LIDC-IDRI and IQ-OTH/NCCD CT image datasets that provided state-of-the-art results.
v. Also, an ablation study is carried out to prove the efficiency and efficacy of the suggested framework.

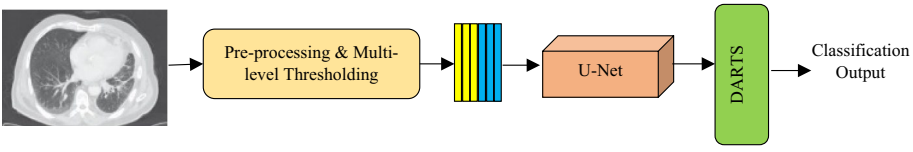


Fig. 1 Block Diagram for the proposed UDCT Framework

The organization of the rest of the paper is as follows: Section 2 presents the different state-of-the-art techniques for lung cancer detection, Section 3 presents the proposed methodology, Section 4 discusses the results obtained on the proposed UDCT framework. Finally, Section 5 concludes the work.

2 Related work

In [1], a thorough analysis was conducted on various CNN-based techniques and algorithms to analyse pulmonary nodules in medical images. The focus was on examining the methodologies employed, the datasets used for validation, and the detection outcomes of recently developed CNN-based systems for nodule classification. Three different architectures, each consisting of varying numbers of layers, were studied, leading to a fully connected layer and a SoftMax classifier. The LIDC-IDRI dataset was utilized for training and testing, as it yielded the best results in terms of accuracy and precision when combined with other architectures [8]. A state-of-the-art performance was achieved for both lung nodule identification and malignancy classification tasks using a cascade of 3D CNNs in a CADe & CADx system. However, if the CADe model fails to detect a malignant lesion, the subsequent CADx model is unable to classify that scan as malignant, resulting in a significant source of false negatives in the system [10].

To find pulmonary nodules in the lung's current CT pictures, a convolution neural network that has been genetically optimised is used [9] and also the detection rate and classification accuracy of pulmonary nodules can be effectively increased with the use of wavelet denoising on lung CT images. [13] You Only Look Once, Version 3 (YOLOv3), Single-Shot Detector (SSD), and Faster Region-Convolutional Neural Network (Faster-RCNN) are components of an automated deep learning system that is utilized for detection and Visual Geometry Group (VGG-16), Resnet 101, and Mobilenet for classifying lung nodules in CT scans. In this study, False Positive Rate (FPR) is reduced, and the accuracy is enhanced for classification and detection architectures. A hybrid approach that further aids in reducing the computational complexity of the classification is produced by combining Support Vector Machine with Feed-Forward Back Propagation Neural Network (FFBPNN) [14]. By using a hybrid classification algorithm, the classification accuracy is enhanced to 98.08%. To perhaps increase classification accuracy in the future, Deep Neural Network are used in contrast to FFBPNN. To identifying and categorizing lung cancer, a comprehensive framework is presented that includes a generative model designed to transform an unbalanced dataset which is smaller in size into a balanced one which is larger in size and ResNet50 which classifying CXR pictures into benign and malignant during the online phase by using a large balanced dataset [11]. In this, authors want to modify their system in the future to incorporate generative adversarial networks and other generative models. [30] Presents an optimization-driven technique for lung cancer classification. Lung cancer is classified using a Hierarchical Attention Network (HAN), Firefly Competitive Swarm Optimisation (FCSO), which combines the Firefly algorithm (FA) and Competitive Swarm Optimisation (CSO). The suggested FCSO-based HAN performs well, achieving a high accuracy rate of 91.3%, sensitivity of 88%, and specificity of 89.1%. The author addresses the convergence problems and computational overhead related to the multilevel thresholding process by utilizing the Levy flight and Gravitational Search Algorithm (LCGSA) based on Chaos theory. This technique is used to improve COVID-19 chest CT scan images

segmentation [31]. The author may also look into the future use of LCGSA in deep learning applications, such as CNNs, U-Nets, and LSTMs.

The authors of this paper addressed the scalability issue in architecture search by utilizing DARTS, a differential approach. This strategy depends on continuous or smooth relaxation, in contrast to traditional approaches that use evolutionary algorithms or reinforcement learning inside a discrete and non-smooth search space. The search was performed on multiple datasets including WikiText-2, CIFAR-10, Penn Treebank, and ImageNet. However, a limitation of this work is the potential discrepancies between the continuous architecture encoding and the resulting discrete architecture, which can impact the effectiveness of the current approach [15].

In another study, blood samples were taken from healthy individuals and those who had been diagnosed with breast cancer were subjected to metabolomics analysis using NMR spectroscopy and direct analysis in real time (DART)-MS. The new method introduced in this research improves the representation of disease classifications by allowing the expression of disease on a continuous scale rather than a binary scale, resulting in more accurate disease detection. However, DART-MS did not capture as much breast cancer-related variation [16]. In the same study, U-DARTS was employed to identify and categorize breast lesions in two datasets, achieving testing and validation accuracies of 91% and 98%, respectively. However, due to incomplete data, the study encountered low accuracy in one of the datasets [17].

Methods discussed above are mainly utilized for detecting and classifying lung cancer that may have certain limitations in a variety of areas, including accuracy, sensitivity, specificity, dataset availability, preprocessing techniques, and sometimes the classification models used. We present a novel UDCT framework with Multi-Level Otsu thresholding to address these issues and to improve image quality during preprocessing. Furthermore, to obtain higher accuracy, we combine U-NET and DARTS to improve lung cancer classification. Furthermore, different methods for the lung cancer detection are presented in Table 1.

3 Proposed methodology

This section describes the proposed UDCT framework (as depicted in Fig. 2) for the detection and classification of lung diseases. The proposed framework utilized a modified U-NET model which exploits DARTS architecture for efficient classification of lung cancer diseases. Initially, to improve the quality of the input images, an image enhancement technique is applied through pre-processing. The pre-processing step involves multilevel Otsu for better results. After pre-processing, the image is segmented using U-NET, and quantitative features are extracted. To enhance the classification accuracy, we fuse DARTS (Differentiable Architecture Search) with the U-NET model. The overall goal of this proposed work is to achieve improved results in the classification of lung diseases, particularly lung cancer.

3.1 Problem statement

In order to differentiate between benign and malignant nodules, computer-aided detection techniques for lung cancer require an examination of the visual properties of lung nodules on CT scans. Developing an automated detection method is becoming more and more crucial for early detection of lung cancer. It is incredibly challenging to build such a system using traditional approaches.

Table 1 Review based on some existing method for lung cancer detection

References	Approach	Evaluated Parameters
[1]	CNN-based techniques	–
[8]	Deep Convolutional Neural Networks (DCNN)	Accuracy = 80.02%, Precision = 84.51%, Recall = 83.45% and Dice Score Coefficient = 81.65%
[10]	Computer-Aided Detection (CADe) and Computer-Aided Diagnosis (CADx)	–
[9]	Wavelet analysis algorithm and Genetic Algorithm	Error = 0.01% which increased accuracy
[13]	Faster-RCNN, YOLOv3, SSD, VGG-16, Resnet 101, and Mobilenet	–
[14]	Feed-Forward Back Propagation Neural Network (FFBPNN)	Accuracy = 98.08%
[11]	Convolutional Variational Generative Auto-encoder and ResNet50	Accuracy = 98.91%, Area Under Curve (AUC) = 98.85%, Sensitivity = 98.46%, Precision = 97.72% and F1 score = 97.89%.
[15]	Differentiable Architecture Search (DARTS)	Search Cost (GPU Days) =4
[16]	NMR spectroscopy and Direct Analysis in real time Mass spectrometry (DART)-MS	AUC = 0.99
[17]	Modified Differentiable Architecture Search(U-DARTS)	Validation Accuracy = 98% and Testing Accuracy = 91%
[31]	Hierarchical Attention Network (HAN), Firefly Competitive Swarm Optimisation (FCSO) which combines the Firefly algorithm (FA) and Competitive Swarm Optimisation (CSO)	Accuracy = 91.3%, Sensitivity = 88%, and Specificity = 89.1%.
[32]	Levy flight and Chaos theory-based Gravitational Search Algorithm (LCGSA)	Accuracy = 92%, Precision = 1, Recall = 83% and Dice Coefficient = 90%

Deep-learning-based architectures have been extensively researched and developed in the last few years with the goal of being used to the healthcare industry. Thus, the goal of this work was to develop a deep learning-based, reliable, and accurate framework for lung cancer detection and classification. Furthermore, the pseudocode for the proposed methodology is presented in Algorithm 1.

Algorithm 1 Pseudocode for the proposed UDCT Framework for Lung Cancer detection and classification for CT Images.

Input: $I_{CT}(xxy)$; Thermal image with x=no. of rows and y= no. of column

Output: I_{UDART}; Output Image

Procedure UDCT

Initialize the image $I_{CT}(xxy)$

Apply Otsu multilevel thresholding (OMLT) on I_{CT} to separate object pixels from background pixels.

$I_{CT-OMLT} \xleftarrow[OMLT]{} I_{CT}$

For Normal (N) & Abnormal (A) class calculate thresholding value X

$N \leftarrow p$ if $0 \leq p < X$,

$A \leftarrow p$ if $X \leq p < L\text{-}1$

Where p is one of pixels of the grayscale image that can be represented in L grayscale levels

Apply UNet segmentation on I_{CT} image.

while $(x < UNet_{\text{iterations}})$ do

for each patch do

Update the image patches obtained through segmentation using eq (14)

end for

end while

Update I_{CT}

for i = 1 to n (n= no. of column in image)

$U(I_{CT}) = U(I_{CT-OMLT})$

end for

Use DARTS for classification of abnormal and normal images.

return archive

end procedure

3.2 Multilevel Thresholding using Otsu method

Image thresholding is a method used to convert an image, whether it is in full-color or grayscale, into a binary image where pixels are either classified as foreground or background. Its main purpose is to separate object pixels from background pixels. The determination of an appropriate threshold value plays a critical role in the segmentation process. Here, threshold value X is calculated as:

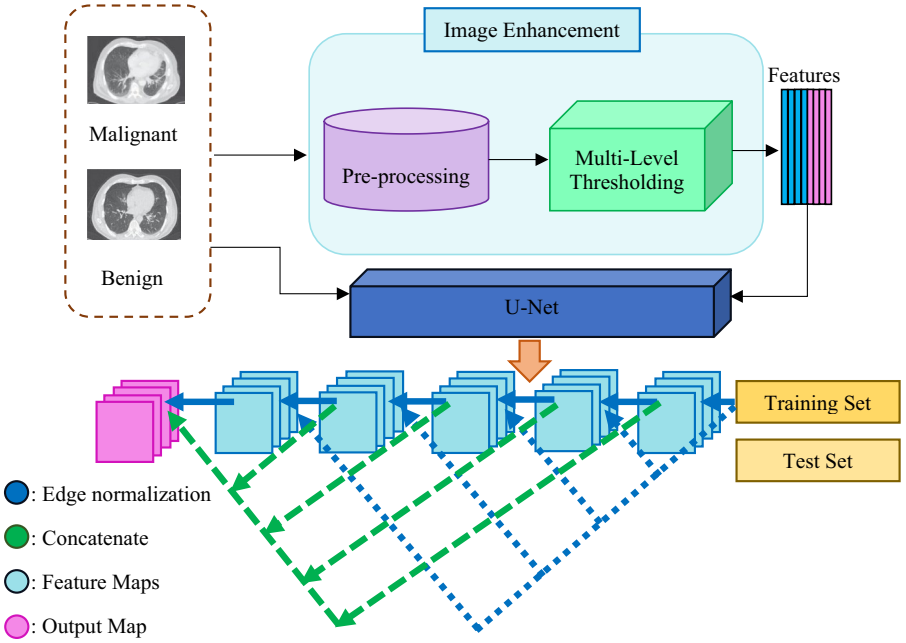


Fig. 2 Proposed UDCT Framework Methodology

$$\mathrm{N} \leftarrow \mathrm{p} \text{ if } 0 \leq \mathrm{p} < \mathrm{X}, \tag{1}$$

$$\mathrm{A} \leftarrow \mathrm{p} \text{ if } \mathrm{X} \leq \mathrm{p} < \mathrm{L} - 1$$

Where, N and A are the Normal & Abnormal classes when X is the threshold and p is one of pixels of the grayscale image that can be represented in L grayscale levels $L = \{0, 1, 2, \ldots, L-1\}$.

Equation (1) relate to the bi-level threshold and expanded to many sets like:

$$\mathrm{N} \leftarrow \mathrm{p} \text{ if } 0 \leq \mathrm{p} < X_1, \tag{2}$$

$$\mathrm{A} \leftarrow \mathrm{p} \text{ if } X_1 \leq \mathrm{p} < X_2,$$

$$N_n \leftarrow \mathrm{p} \text{ if } X_n \leq \mathrm{p} < X_{n+1},$$

$$A_m \leftarrow \mathrm{p} \text{ if } X_m \leq \mathrm{p} < \mathrm{L} - 1$$

In eq. 2, $\{X_1, X_2, \ldots.., X_n, X_{n+1}\}$ are different threshold levels. To determine the ideal threshold value, Otsu proposes various objective functions that must be maximised. For figuring out threshold values, Otsu is a well-known method. The Otsu method, which uses the greatest variance values of several classes as the only segmentation criterion in images, is preferred because of its ease of use and great efficacy. Applying the L designated intensity level for the grayscale image, the intensity value of image computation can be written as [18].

$$\mathrm{P}q_a^o = \frac{q_a^o}{P_{tot}}, \tag{3}$$

$$q_a^o \geq 0, \tag{4}$$

$$\sum_{a=0}^{P_{tot}} Pq_a^o = 1,$$

$$\mathrm{o} = \{1, \text{if Grayscale Image},$$

o = component, the value specified from I.

where,

$\mathrm{P}q_a^o$ = Distributional probability.

$\mathrm{P}_{\mathrm{tot}}$ = the total number of pixels in the image.

q_a^o =histogram (the number of pixels correlates to the intensity in o).

Now, Probability distributions for two classes (N and A) segmentation are given as [18]:

$$H_0^o(\mathrm{X}) = \sum_{a=0}^{X} Pq_a^o, H_1^o(\mathrm{X}) = \sum_{a=X+1}^{L} Pq_a^o \tag{5}$$

The mean and the variance is being calculated as representation mathematically in equation '6' and '9':

$$\mu_0^o = \sum_{a=0}^{X} \frac{aPq_a^o}{H_0^o(X)} \text{ and } \mu_1^o = \sum_{a=X+1}^{L} \frac{aPq_a^o}{H_1^o(X)} \tag{6}$$

$$\text{and } \mu_{tot}^o = H_0^o \mu_0^o + H_1^o \mu_1^o, \tag{7}$$

$$H_0^o + H_1^o = 1 \tag{8}$$

The total variation between the classes is computed after the calculation of the above values as,

$$\sigma^{2^o} = \sigma_1^o + \sigma_2^o \tag{9}$$

In the mathematical sense, number two does not represent exponent; it is a component of the Otsu variant operator in eq. (8). In the above, σ_1^o and σ_2^o are the variants of N and A which are obtained as:

$$\sigma_1^o = H_0^o\left(\mu_0^o + \mu_{tot}^o\right)^2 \& \sigma_2^o = H_1^o\left(\mu_1^o + \mu_{tot}^o\right)^2 \tag{10}$$

In (10), the value of $'\mu_{tot}^o{}'$ is find with the help of eq. (8).

For multilevel thresholding, the objective function can be defined by (9) because by the increase of number of thresholds there will be an exponential growth in processing time.

$$\theta(X) = \max\left(\left(\sigma^{2^o}(X)\right), 0 \leq \mathrm{X} \leq \mathrm{L} - 1\right. \tag{11}$$

Where, $\mathrm{X} = \{X_1, X_2, X_3 \ldots\ldots X_{k-1}\}$ there are multiple thresholds in the vector X. So, for various number of threshold values Variance, Probability Distribution and the Average values can be calculated by eqs. (11), (12) and (13) respectively.

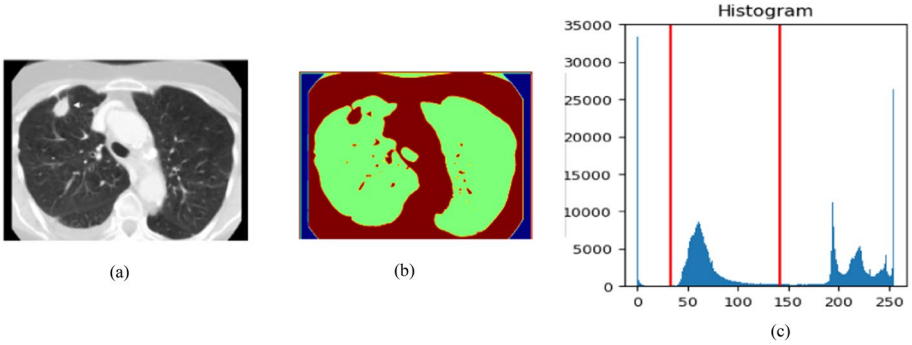


Fig. 3 Multilevel thresholding by using Otsu method, (**a**) Original lung image, (**b**) Result after Otsu thresholding, and (**c**) Equalized Histogram

$$\sigma^{2^o} = \sum_{c=0}^{k} H_c^o \left(\mu_c^o - \mu_{tot}^o\right) \tag{12}$$

$$H_0^o(X) = \sum_{c=1}^{X_1} Pq_c^o, H_{k-1}^o(X) = \sum_{c=X_{K+1}}^{L} Pq_c^o \tag{13}$$

$$\mu_0^o = \sum_{c=1}^{X_1} \frac{cPq_c^o}{H_0^o(X_1)}, H_{k-1}^o = \sum_{c=X_{k+1}}^{L} \frac{cPq_c^o}{H_1^o(X_k)} \tag{14}$$

Where, 'c' represents a class in the above equations.

Furthermore, the output obtained after multilevel thresholding on LIDC dataset is presented in Fig. 3. Multilevel thresholding is a technique that partitions an image containing various shades of gray into multiple distinct regions. This approach involves choosing multiple thresholds to separate the image into different brightness zones, representing various backgrounds and objects.

3.3 Image segmentation using U-net

The objective of segmentation is to separate individual elements from an entire image. Segmentation is a procedure that transforms original images into distinct regions or components. The accuracy of the resulting segmentation is used as a measure of how well the process was performed. The precision of the segmentation increases when the objects are recognized more accurately. In the field of biomedical imaging, the U-Net architecture is commonly utilized for image segmentation tasks. This architecture consists of two parts: a contracting side and an expanding side, which are symmetrical. The contracting side includes components such as dropout, pooling, convolutional, and Rectified Linear Unit (ReLU) layer. Each down sampling step in the process doubles the number of feature channels [19].

i. Down-sampling is performed in contracting path with the shallower, lower, and coarser grained characteristics are converted into the deeper, semantic, and finer grained features by this connection [20].
ii. The connection between contracting and expanding sides called skip connection enable the network to take in the required features from the relevant layer. Skip con-

nection bridges the shallower, lower, and coarse-grained features in the contracting path and the deeper, semantic, and fine-grained features in the expanding path.

iii. Up-sampling is performed in expanding path which transforms back Deeper, semantic, and fine-grained characteristics into shallower, lower, and coarse-grained features. At every section on the expanding side, the number of feature channels is cut in half [20].

Using the channel concatenation fusion approach, let $Z^{u,v}$ represent the $v-th$ convolution in the $u-th$ layer module. The output $z^{u,v}$ of $Z^{u,v}$can be calculated according to the following eq. [20]:

$$z^{u,v} = \begin{cases} \mathrm{M}\left(\mathrm{D}\left(z^{u-1,v}\right)\right) & \text{for u} > 0, \text{v} = 0 \\ \mathrm{M}\left(\left[\left[z^{u,l}\right]_{l=0}^{v-1}, \mathrm{S}\left(z^{u+1,v-1}\right)\right]\right) & \text{for u} = 0, \text{v} > 0 \\ \mathrm{M}\left(\left[\left[z^{u,l}\right]_{l=0}^{v-1}, \mathrm{S}\left(z^{u+1,v-1}\right), \mathrm{D}\left(z^{u-1,v}\right)\right]\right) & \text{for u} > 0, \text{v} > 0 \end{cases} \tag{15}$$

In the above equation, M(•) is a cascaded operation including convolution, ReLU activation function and batch normalization (BN), D(•) is a down- sampling function, S(•) is an up-sampling function, and [•] is the channel concatenation operation [20].

After multi-level thresholding, we have extracted following qualitative features with the following value: *Mean* = 1.925, *Standard Deviation* = 59.283, *Dissimilarity* = 0.6221, *Energy* = 24.171 *and Entropy* = 8.177 for LIDC-IDRI dataset. Furthermore, for IQ-OTH/NCCD dataset, *Mean* = 0.916, *Standard Deviation* = 50.302, *Dissimilarity* = 01.462, *Energy* = 23.059 *and Entropy* = 8.257 The different features shown with the help of Fig. 4. Along with the above network we explained DARTS (Differentiable Architecture Search) for searching the efficient architecture. Darts has had a significant impact on neural architecture search. The representation of DARTS is presented in Fig. 5. Earlier techniques relied on reinforcement learning and needed plenty of processing power. It required 3150 GPU days for evolution or 2000 Graphics Processing Unit (GPU) days for reinforcement learning [15].

We relax the search space to be continuous so that the architecture can be optimised via gradient descent with regard to the performance of the validation set. This allows us to search over a continuous universe of candidate architectures rather than a discrete set. This search is based on bilevel optimization and also relevant to recurrent structures as well as convolutional architectures. In the search space, a directed acyclic graph depicts the computation process for the proposed architecture as depicted in Fig. 5. A Directed acyclic graph with F nodes ($f_0, f_1, f_2 \ldots\ldots f_F$) arranged in an orderly manner.

In this work, we take B be a collection of operations which includes convolution and max pooling, where each operation denotes a function to be applied on $r(t)$ as $b(.)$. Here all possible operations are given by $b^{(g,h)}$ [15]:

$$b^{(g,h)}(r) = \sum\nolimits_{b \epsilon B} \frac{\exp\left(y_b^{(g,h)}\right)}{\sum_{b' \epsilon B} \exp\left(\gamma_{b'}^{(g,h)}\right)} r(t) \tag{16}$$

For a pair of nodes (g, h) a vector $\gamma^{(g,h)}$ of dimension $|B|$ is used as a parameter for the operation of mixing weights .After that, learning a set of continuous variables is all that

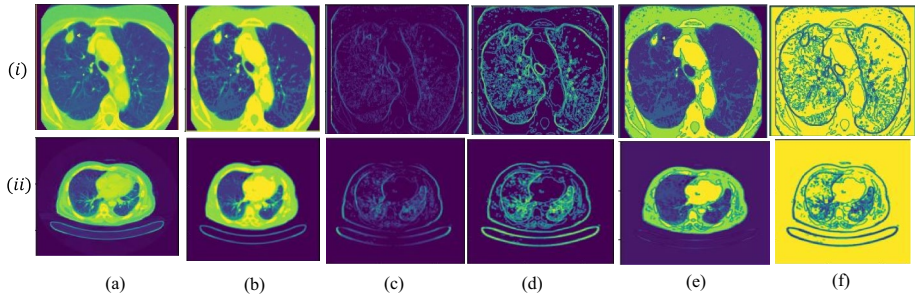


Fig. 4 Qualitative features extracted from the proposed U-NET approach for (*i*) LIDC-IDRI dataset, (*ii*) IQ-OTH/NCCD dataset. (**a**) Original source image from dataset, (**b**) mean value, (**c**) Dissimilarity, (**d**) Entropy, (**e**) Standard Deviation, and (**f**) Energy

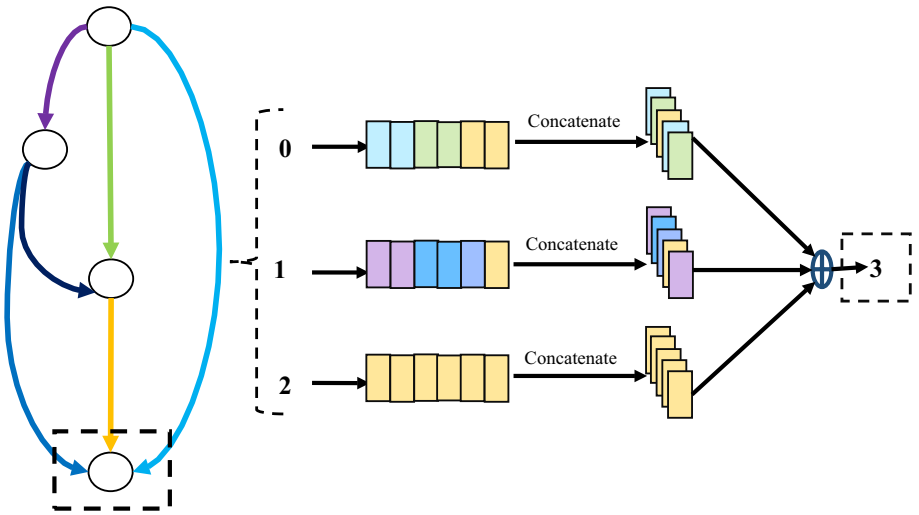


Fig. 5 Representation of DARTS

remains of the architectural search task $\gamma = (\gamma^{(g,h)})$ [17]. Based on each predecessor, each intermediate node is calculated by 'r(t)' mentioned in eq. 17 as follows:

$$\mathrm{r(t)} = \sum_{g<h} b^{(g,h)}(r(t)) \tag{17}$$

In this work, we have fused DARTS with UNET for getting better lung cancer classification results.

3.4 Dataset used

To assess the effectiveness of the proposed UDCT framework for the categorization of benign and malignant pulmonary nodules, the LIDC-IDRI public dataset of thoracic CT scans and IQ-OTH/NCCD dataset are used.

LIDC-IDRI stands for Lung Image Database Consortium and Image Database Resource Initiative. On a CT scan, lung nodules can be seen in several slices. Each case includes a clinical thoracic CT scan as well as the reports of four expert thoracic radiologists in an XML file [13]. Additionally, the XML files include semantic diagnostic elements that were noted by four qualified thoracic radiologists. The Lung Imaging Database Consortium (LIDC) initiative's goal is to assist a group of institutions in creating a database of spiral CT lung pictures and developing recommendations for a resource for spiral CT lung images. There are 1568 thoracic Computed Tomography (CT) scans with nodules in the dataset out of which 650 were malignant cases and 918 were benign cases [19]. There are three different types of nodules: ("nodule > or =3 mm," "nodule <3 mm," and "non-nodule > or =3 mm") [21].

The IQ-OTH/NCCD dataset, which stands for Iraq-Oncology Teaching Hospital/ National Center for Cancer Diseases, includes CT scans of both healthy persons and those with various stages of lung cancer diagnosis. It has 1190 images total, each of which shows slices from 110 CT scans. Three classes—normal, benign, and malignant—are used to group these situations. 40 cases are categorized as malignant, 15 as benign, and 55 as normal. DICOM is the original format used to store CT scan data.

4 Computational complexity

The computational complexity of the proposed UDCT framework algorithm is described in this subsection. The following describes the space and time complexity:

- **Time Complexity**

1. The implementation of UDCT algorithm desires $O(I_{CT}(x \times y))$ time where x and y represent the number of rows columns respectively in each image.
2. The OMLT calculation of each pixel requires

$$O\left(I_{CT}(OMLT \times x \times y)\right)$$

time.

3. The UNet calculations require

$$O\left(I_{CT}\left(U_{\text{iter}} \times OMLT \times x \times y\right)\right)$$

time where U_{iter} is the maximum number of iterations to simulate segmentation.

4. The algorithm requires $O(I_{UDART}) = O(I_{CT-DART}(O(I_{CT}(U_{\text{iter}} \times OMLT \times x \times y)))$ $I_{CT-DART}(O(I_{CT}(U_{\text{iter}} \times OMLT \times x \times y)))$ time to classify abnormal and abnormal type of cancer.

- **Space Complexity**

The UDCT algorithm's space complexity is determined at the initial stage, which at any given time requires space. Due to this, the UDCT algorithm's overall space complexity is $O(I_{CT}(x \times y))$.

5 Results and discussions

This section discusses the implementation details, and evaluation parameters with the experimental results obtained on LIDC-IDRI and IQ-OTH/NCCD CT image datasets.

5.1 Implementation details

We train our model from beginning to end in order to identify and extract image features from the lung images. The patch size is set to 9 and the batch size is set as 32. The input images are downsized to 124×124 resolution. A learning rate 0.0001 is used to train the entire model on the dataset. For the setup, there will be 100 epochs. Accuracy and loss curves are plotted to evaluate the effectiveness of the proposed model. Further, ThinkPad laptop that satisfies certain requirements is used for the experiment. It has an Intel Core i7 (11th generation) processor with integrated AMD Radeon Graphics, clock rates ranging from 1.90 GHz to 4.40 GHz, and 16 GB of RAM. The study's experiments were conducted on Windows 11 with Jupyter Notebook, Python 3.12.3, and the backend libraries for TensorFlow and Keras.

5.2 Evaluation parameters

Accuracy, Loss, Precision, Recall, Dice Coefficient, Specificity and F1-Score are the several parameters that are used to evaluate the proposed model's performance.

i) **Accuracy**: It is defined by the proportion of accurate predictions to all predictions [22]. Represented by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

ii) **Recall**: Ratio of the number of Positive samples that were correctly identified as Positive to all Positive samples. More positive samples are detected when the value of recall increases [22].

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

iii) **Precision**: It shows the percentage of samples that were actually positive and those that were projected to be positive out of all the samples [22].

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

iv) **Dice Coefficient**: It is calculated by dividing the combined pixels in the two images by the area of overlap, multiplied by 2.

$$DSC = \frac{2 * |D \cap E|}{|D| + |E|} \tag{21}$$

Where D is the expected set of pixels and E is the actual set of pixels.

xxii) Specificity: The ratio of real negative samples—both actual and predicted—to the total number of truly negative samples is used to compute specificity [22].

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{22}$$

vi) F1-Score: The "F1-score" accounts for both false positives (FPs) and false negatives (FNs), striking a compromise between "precision" and "recall". Since each variable undermines the other, it penalizes extreme levels of "precision" and "recall" [22].

$$\text{F1} = \frac{2 * P * R}{P + R} \tag{23}$$

Where P and R represents Precision and Recall Respectively.

In the eqs. (18) to (22), *TP*, *TN*, *FN*, *and FP* represents true positive, true negative, false negative, and false positive, respectively.

5.3 Performance analysis on LIDC-IDRI and IQ-OTH/NCCD CT image datasets

The experimental findings for the proposed U-DARTS model are detailed in this section. The LIDC-IDRI dataset is used for the experiment. The sample contains a wide variety of lung nodules, which are separated into two groups: Benign (Unharmful) and Malignant (harmful). The experiment is done by using 80% training set and 20% testing set. Training and validation accuracy, as well as training and validation loss, are used to assess the proposed framework using the LIDC-IDRI dataset. Figures 6 and 7 shows training-validation loss and training-validation accuracy respectively for LIDC-IDRI and IQ-OTH/NCCD datasets. From these accuracy and loss curves it is evident that the as the number of epochs increases the accuracy increases whereas the loss decreases. Furthermore, the training and validation accuracy obtained for the proposed UDCT framework on LIDC-IDRI and IQ-OTH/NCCD is 93.5% and 95.01% and 94.8% 96.82% respectively. Furthermore, the training and validation losses attained are 0.32 and 0.28 for LIDC dataset and 0.38 and 0.25 respectively on IQ-OTH/NCCD.

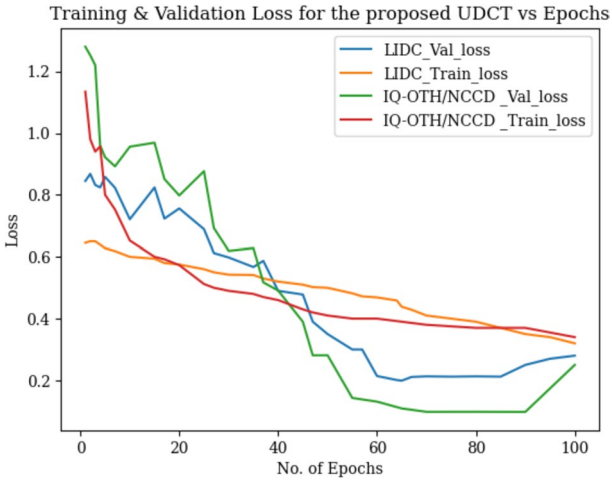


Fig. 6 Validation and Training Loss for the proposed UDCT Framework for LIDC and IQ-OTH Dataset

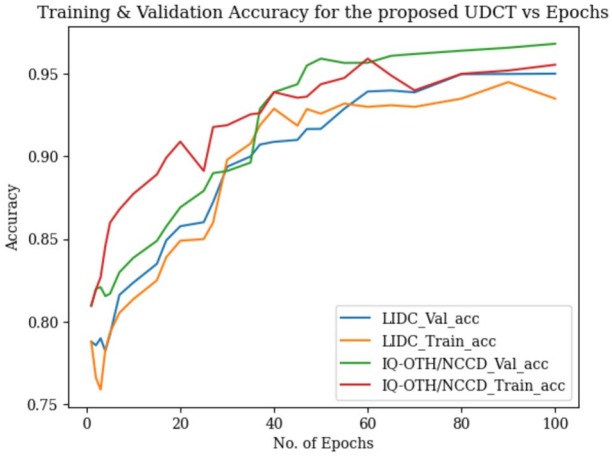


Fig. 7 Validation and Training Accuracy for the proposed UDCT Framework for LIDC and IQ-OTH Dataset

Tables 2 and 3 reports the comparison results obtained for the proposed UDCT framework with other state-of-the-art on LIDC-IDRI and IQ-OTH/NCCD CT image datasets for lung cancer detection and classification. The proposed work achieved an accuracy of 95.01%, Validation Loss of 0.28, Precision of 97.00%, Recall of 97.52%, Dice coefficient of 96.10%, Specificity of 99.09% and F1-Score of 96.01% on LIDC-IDRI Dataset. This proves that U-DARTS produces greater accuracy than the most effective approaches [24–29] presented in the Table 2. The proposed model offers relatively high accuracy when compared to the previous work because of these model adjustments, the usage of segmentation techniques, and the larger datasets as a result. Furthermore, a comparison for the proposed framework is made for the IQ-OTH/NCCD Dataset as reported in Table 3 which provided state-of-the-art results when compared with [33–38]. Also, Fig. 8(a) and (b) presents the confusion matrix for the proposed UDCT framework on LIDC and IQ-OOTH/NCCD datasets.

5.4 Ablation study

(i) An Ablation study is conducted to validate the performance of the proposed UDCT framework. In this section, we have performed three shuffles of the data, as depicted in Table 4, using fold 1, fold 2, and fold 3. This table illustrates how different evalu-

Table 2 Performance comparison with previous work done on LIDC-IDRI dataset

Model	Accuracy	Loss	Precision	Recall	Dice coefficient	Specificity	F1-Score
Wu et al. [23]	–	–	–	97.3%	80.2%	–	–
Yu et al. [24]	87.3%	–	77.1%	–	69.8%	–	–
Banu et al. [25]	94.66%	–	–	90.84%	90.35%	–	–
Bruntha et al. [26]	–	–	89.09%	81.08%	87.23%	–	–
Bruntha et al. [27]	–	–	96.75%	92.57%	95.7%	–	–
Bushara et al. [28]	94%	–	95%	94.5%	–	99.07%	94.5%
Shafi et al. [29]	94%	–	95%	94.5%	–	–	94.5%
UDCT	**95.01%**	**0.28**	**97.00%**	**97.52%**	**96.10%**	**99.09%**	**96.01%**

Table 3 Performance comparison with previous work done on *IQ-OTH/NCCD* dataset

Model	Accuracy	Loss	Precision	Recall	Dice coefficient	Specificity	F1-Score
Huseiny et al. [32]	94.38%	–	–	95.08%	–	93.70%	–
Mathews et al. [33]	95.00%	–	–	97.00%	–	98.00%	–
Suresha et al. [34]	–	–	98.40%	97.00%	–	–	97.70%
Al-Yasriy et al. [35]	93.54%	–	97.10%	95.71%	–	95.00%	96.40%
Gowda et al. [36]	96.50%	–	–	89.00%	–	94.00%	–
Santos et al. [37]	95.80%	–	–	99.20%	–	98.30%	–
Kareem et al. [38]	89.88%	–	98.55%	97.14%	–	97.50%	97.84%
UDCT	**96.82%**	**0.25**	**98.70%**	**97.50%**	**96.4%**	**98.40%**	**98.24%**

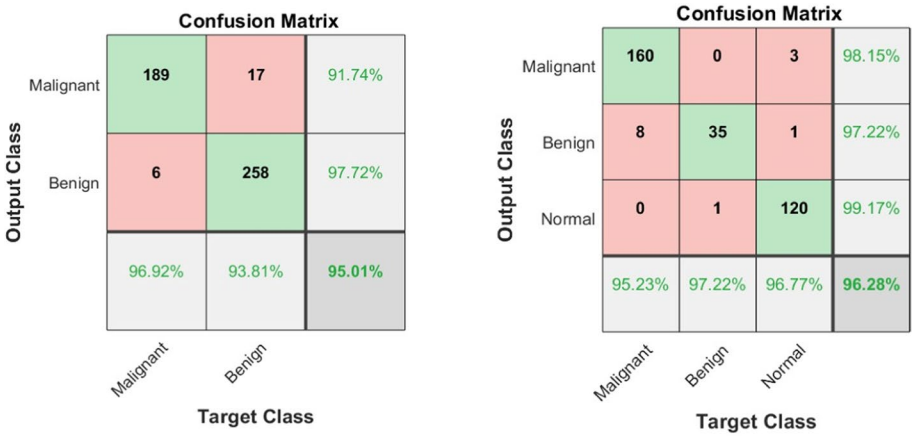


Fig. 8 Confusion Matrix for the proposed UDCT framework on (**a**) LIDC Dataset and (**b**) IQ-OTH Dataset

ation parameters vary across the different folds. The effectiveness of the method is demonstrated by employing a three-fold cross-validation approach in this study. Each fold of the cross-validation process allows us to assess the impact of the training samples on the accuracy, precision, recall, and Dice coefficient of the classification results by considering a specific fraction of the training data. Also, a comparison graph is depicted in Fig. 9 for three different folds with two classes each for one-fold.

(ii) An ablation is also conducted to study the influence batch size with optimizer utilized and learning rate. The results for the same are reported in Table 5. From these results we can infer that the best accuracy for 95.01% is obtained on Adam optimizer with a learning rate of 0.0001 for batch size of 32. Based on our results, we can say that the batch size and learning rate significantly affect the performance of the framework. Also, this experimental study aims at providing a better understanding of the batch size value which is 32 with the best suited optimizer and learning rate for the classification problems

Table 4 Ablation Study Results Obtained for the proposed UDCT Framework on LIDC-IDRI dataset

Fold	Classes	#Samples	Performance Measure			
			Accuracy	Precision	Recall	Dice Coefficient
Fold 1	Benign	918	0.909	0.912	0.873	0.901
	Malignant	650	**0.950**	0.961	0.966	0.948
Fold 2	Benign	918	0.910	0.904	0.908	0.890
	Malignant	650	0.928	0.950	0.951	0.922
Fold 3	Benign	918	0.904	0.935	0.870	0.907
	Malignant	650	0.948	**0.970**	**0.975**	**0.961**

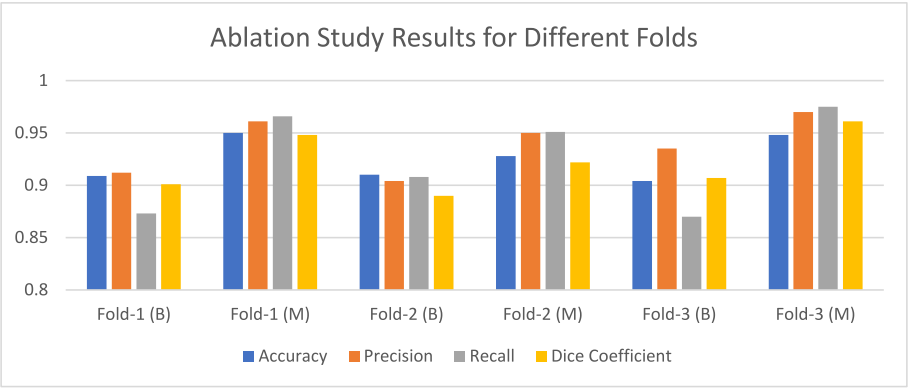


Fig. 9 Ablation Study Comparison Results for LIDC Dataset

Table 5 Ablation study to study the influence of different hyperparameters on the proposed UDCT Framework on LIDC Dataset

Batch Size	Optimizer			
	Adam		SGD	
	LR = 0.001	LR = 0.0001	LR = 0.001	LR = 0.0001
16	93.71%	92.2%	87.55%	89.64
32	94.21%	95.01%	89.90%	92.32
64	93.35%	94.00%	89.92%	92.35
128	93.87%	93.20%	91.23%	92.80

6 Conclusion

This paper presents a novel framework, UDCT for automatic detection of nodules on CT images. The proposed framework utilizes Multilevel Otsu thresholding for accurate segmentation of lung tumours from CT scans. Multilevel Otsu thresholding helps to separate lung tissue from the CT scans. Further, the proposed model incorporates U-Net network with robust multi-scale feature extraction capabilities. Also, DARTS-based classifier is incorporated which utilizes the concept of continuous relaxation of the architecture representation with efficient gradient descent search for efficient classification of lung cancer.

The proposed framework combines the knowledge of DARTS and U-Net with aims to improve lung nodule classification into benign and malignant classes. The proposed technique was evaluated and validated on the publicly available LIDC-IDRI dataset and IQ-OOTH/NCCD dataset and provided state-of-the-art results. The results reported provided an accuracy of 95.01%, a dice coefficient of 96.10%, and a recall rate of 97.52% on LIDC-IDRI dataset. Also, accuracy of 96.82%, a dice coefficient of 96.40%, and a recall rate of 97.50% on IQ-OOTH/NCCD dataset. Further, an ablation study is also conducted to validate the efficiency and efficacy of the proposed model. In future, we can extend our experiment on large scale dataset. Also, weak supervised learning algorithm can be incorporated that utilizes large number of incomplete, inaccurate, and ambiguous annotation data in the existing medical records to achieve efficient model training. Furthermore, bringing prior clinical knowledge into model training should be encouraged so as the model provides the best possible classification output.

Funding No funding was received to assist with the preparation of this manuscript.

Data availability The datasets generated during and/or analyzed during the current study are available through an online web repository via https://wiki.cancerimagingarchive.net/display/Public/LIDC-IDRI and https://www.kaggle.com/datasets/adityamahimkar/iqothnccd-lung-cancer-dataset.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Monkam P, Qi S, Ma H, Gao W, Yao Y, Qian W (2019) Detection and classification of pulmonary nodules using convolutional neural networks: a survey. IEEE Access 7:78075–78091. https://doi.org/10.1109/ACCESS.2019.2920980
2. Kumar V, Bakariya B (2021) Classification of malignant lung cancer using deep learning. J Med Eng Technol 45(2):85–93. https://doi.org/10.1080/03091902.2020.1853837
3. Fontana RS et al (1984) Early lung cancer detection: results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic study. Am Rev Respir Dis 130(4):561–565. https://doi.org/10.1164/arrd.1984.130.4.549
4. Blandin Knight S, Crosbie PA, Balata H, Chudziak J, Hussell T, Dive C (2017) Progress and prospects of early detection in lung cancer. Open Biol 7(9):170070. https://doi.org/10.1098/rsob.170070
5. Inage T, Nakajima T, Yoshino I, Yasufuku K (2018) Early lung Cancer detection. Clin Chest Med 39(1):45–55. https://doi.org/10.1016/j.ccm.2017.10.003
6. D. S, B. R, M. T, P. P, D. T, and V. S (2022) Detection and Stage Classification of UNet Segmented Lung Nodules Using CNN," in 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT). 1–5. https://doi.org/10.1109/IMPACT55510.2022.10029182
7. Manser R et al (2013) Screening for lung cancer (review) screening for lung cancer. Cochrane Database Syst Rev (6). http://www.thecochranelibrary.com
8. Mohanapriya N, Kalaavathi B, senthil Kuamr T (2019) Lung Tumor Classification and Detection from CT Scan Images using Deep Convolutional Neural Networks (DCNN), in 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) pp. 800–805. https://doi.org/10.1109/ICCIKE47802.2019.9004247
9. Li G et al (2020) Study on the detection of pulmonary nodules in CT images based on deep learning. IEEE Access 8:67300–67309. https://doi.org/10.1109/ACCESS.2020.2984381
10. Ozdemir O, Russell RL, Berlin AA (2020) A 3D probabilistic deep learning system for detection and diagnosis of lung Cancer using low-dose CT scans. IEEE Trans Med Imaging 39(5):1419–1429. https://doi.org/10.1109/TMI.2019.2947595

11. Salama WM, Shokry A, Aly MH (2022) A generalized framework for lung Cancer classification based on deep generative models. Multimed Tools Appl 81(23):32705–32722. https://doi.org/10.1007/s11042-022-13005-9
12. Schneider J (2006) Tumor Markers in Detection of Lung Cancer pp. 1–41
13. Ahmed I, Chehri A, Jeon G, Piccialli F (2022) Automated pulmonary nodule classification and detection using deep learning architectures. IEEE/ACM Trans Comput Biol Bioinform:1–12. https://doi.org/10.1109/TCBB.2022.3192139
14. Nanglia P, Kumar S, Mahajan AN, Singh P, Rathee D (2021) A hybrid algorithm for lung cancer classification using SVM and neural networks. ICT Express 7(3):335–341. https://doi.org/10.1016/j.icte.2020.06.007
15. Liu H, Simonyan K, Yang Y (2018) DARTS: Differentiable Architecture Search. http://arxiv.org/abs/1806.09055
16. Gu H, Pan Z, Xi B, Asiago V, Musselman B, Raftery D (2011) Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: application to the detection of breast cancer. Anal Chim Acta 686(1–2):57–63. https://doi.org/10.1016/j.aca.2010.11.040
17. Rautela K, Kumar D, Kumar V (2022) Dual-modality synthetic mammogram construction for breast lesion detection using U-DARTS. Biocybern Biomed Eng 42(3):1041–1050. https://doi.org/10.1016/j.bbe.2022.08.002
18. Nasir M, Farid MS, Suhail Z, Khan MH (2023) Optimal Thresholding for multi-window computed tomography (CT) to predict lung Cancer. Appl Sci 13(12):7256. https://doi.org/10.3390/app13127256
19. Tang S, Ma R, Li Q, Bai Y, Chen S (2021) Classification of benign and malignant pulmonary nodules based on the multiresolution 3D DPSECN model and Semisupervised clustering. IEEE Access 9:43397–43410. https://doi.org/10.1109/ACCESS.2021.3060178
20. Cao Y et al (2023) Segmentation of lung cancer-caused metastatic lesions in bone scan images using self-defined model with deep supervision. Biomed Signal Process Control 79:104068. https://doi.org/10.1016/j.bspc.2022.104068
21. Niranjan Kumar S et al. (2021) Lung Nodule Segmentation Using UNet, in 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) 420–424. https://doi.org/10.1109/ICACCS51430.2021.9441977
22. Ibrahim DM, Elshennawy NM, Sarhan AM (2021) Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases. Comput Biol Med 132:104348. https://doi.org/10.1016/j.compbiomed.2021.104348
23. Wu W, Gao L, Duan H, Huang G, Ye X, Nie S (2020) Segmentation of pulmonary nodules in CT images based on 3D-UNET combined with three-dimensional conditional random field optimization. Med Phys 47(9):4054–4063. https://doi.org/10.1002/mp.14248
24. Yu H, Li J, Zhang L, Cao Y, Yu X, Sun J (2021) Design of lung nodules segmentation and recognition algorithm based on deep learning. BMC Bioinform 22(S5):314. https://doi.org/10.1186/s12859-021-04234-0
25. Banu SF, Sarker MMK, Abdel-Nasser M, Puig D, Raswan HA (2021) AWEU-net: an attention-aware weight excitation U-net for lung nodule segmentation. Appl Sci 11(21):10132. https://doi.org/10.3390/app112110132
26. Bruntha PM, Dhanasekar S, Ahmed LJ, Govindaraj V, Pandian SIA, Abraham SS (2023) "Lung_RUNET: a Segmentation Framework for Lung Nodules," in 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) pp. 658–661. https://doi.org/10.1109/ICACCS57279.2023.10113093
27. Bruntha PM, Pandian SIA, Sagayam KM, Bandopadhyay S, Pomplun M, Dang H (2022) Lung_PAYNet: a pyramidal attention based deep learning network for lung nodule segmentation. Sci Rep 12(1):20330. https://doi.org/10.1038/s41598-022-24900-4
28. Bushara AR, Vinod Kumar RS, Kumar SS (2023) LCD-capsule network for the detection and classification of lung cancer on computed tomography images. Multimed Tools Appl 82(24):37573–37592
29. Shafi Imran et al. (2022) An effective method for lung cancer diagnosis from ct scan using deep learning-based support vector network. Cancers 14(21):5457
30. Spoorthi B, Mahesh S (2023) Firefly competitive swarm optimization based hierarchical attention network for lung cancer detection. Int J Image Graph 23(02):2350017
31. Rather SA, Das S (2023) Levy flight and chaos theory-based gravitational search algorithm for image segmentation. Mathematics 11(18):3913
32. Al-Huseiny MS, Sajit AS (2021) Transfer learning with GoogLeNet for detection of lung cancer. Indones J Electr Eng Comput Sci 22(2):1078–1086
33. Mathews AB (2022) Lung Cancer Segmentation and Classification Using Integration of Convolutional Neural Network & Unet Network Over CT Images: A Deep Learning Approach. Int J Manag Technol Soc Sci 7(1):520–534

34. Prabhu, Akshatha. "Lung Cancer Recognition: A Fusion Approach"
35. Al-Yasriy Hamdalla F et al. (2020) Diagnosis of lung cancer based on CT scans using CNN. IOP Conf Ser Mater Sci Eng. Vol. 928. No. 2. IOP Publishing
36. Gowda S, Jayachandran A (2022) Triple SVM integrated with enhanced random region segmentation for classification of lung tumors. Int J Adv Comput Sci Appl 13(10)
37. Santos DF (2023) Tackling Lung Cancer: Advanced Image Analysis and Deep Learning for Early Detection. Authorea Preprints
38. Kareem HF et al (2021) Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. Indones J Electr Eng Comput Sci 21(3):1731

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Proceedings of the 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024)
IEEE Xplore Part Number: CFP24BC3-ART; ISBN: 979-8-3503-7519-0

2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 979-8-3503-7519-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAAIC60222.2024.10575794

Utilising NLP for Enhanced Clinical Text Mining

Divyanshu Katyan
Dept. of Comp. Science & Engineering
Delhi Technological University
Delhi, India
dkdivyanshukatyan@gmail.com

Gaurav Gulati
Dept. of Comp. Science & Engineering
Delhi Technological University
Delhi, India
gauravg9643@gmail.com

Gaurav Upreti
Dept. of Comp. Science & Engineering
Delhi Technological University
Delhi, India
gauravupreti725@gmail.com

***Abstract*—In recent years, Artificial Intelligence (AI) has emerged as a powerful force, driving innovation across various industries. In the realm of healthcare, AI, particularly Natural Language Processing (NLP), is rapidly gaining prominence as a transformative tool with the potential to reshape patient care. Natural Language Processing, a subfield of AI, has proven to be a game changer in healthcare. It plays a pivotal role in improving patient care, empowering healthcare providers, and optimizing decision-making processes.In this research project, key focus is delving into the dynamic landscape of NLP for Clinical Text Mining, highlighting its profound impact on the healthcare industry. Multiple papers have already been written on this topic. These papers focus on a lot of different topics like cancer terminologies, EHR analysis for job extraction, drug attributes etc. These papers include different databases and mention the precision of all the results as well. NLP empowers healthcare providers in several ways. It discusses how NLP streamlines healthcare workflows, leading to more precise and timely diagnosis, personalized treatment plans, and cost-effective healthcare services.The potential of NLP in healthcare is boundless, promising a future where clinical text mining is integral to delivering superior patient care. This project aims to provide an in- depth analysis of NLP's transformative role, its current status and its evolving prospects within healthcare.**

***Index Terms*—Natural Language Processing, Clinical Text Mining, Clinical Text Analysis, Ensemble Learning, Information Extraction, and electronic health records.**

I. Introduction

A. Motivations and Contributions

The Electronic Health Records (EHRs) which are patients' health records stored in electronic have become a cornerstone in the advancement of healthcare delivery as through the aggregation and analysis of patient data, EHRs facilitate evidence-based decision-making, enabling personalised and timely medical interventions that significantly contribute to the improvement of healthcare quality and patient safety. Despite of this advancement, they are not being effectively being used by doctors due to their abundance and lack of time doctors have to read them.

An effective solution that one can think of is applying NLP to this problem and coming up with a solution that can help millions. However application of NLP to this problem is not easy due to several reasons: (1) records formats issue, (2) medical terminologies abbreviations, (3) clinical text exhibits fragmentation and employs shorthand terminology, (4) marked datasets are rare. Hence the paper has described a method which has huge and far reaching applications in the medical domain. Given medical records the models will give the terms that are relevant to the doctors. The applications include: (1) Early detection of health issues through Family Health (FH) records, (2) Pre-treatment identification of allergies to ensure safe administration of anaesthesia and other medical interventions, (3) Streamlining processes to optimise time efficiency for healthcare professionals.

There were several approaches for solving this problem through NLP. The paper focuses on testing out each one of them and finding out the results, improving the efficiency of the model one at a time. Datasets that have been used to train the models were very hard to find mainly due to the privacy issues in medical datasets. The pipeline used to come to the result includes pre-processing of the medical data (descriptions), since there were a lot of irrelevant terminologies that accompanies the informative ones. This process included performing stemming, tokenisation, TF-IDF etc.
Text similarity measures like Levenshtein distance, Cosine similarity and Jaro-Wrinkler distance were used to remove very similar terms. Thereafter several classifiers were used to figure out the result. These include one vs rest classifier, SGD (Stochastic Gradient Descent) classifier and Multilevel Neural Network. Inspired by the wisdom of the crowd principle, encapsulated ensemble technique was used to further take the advantages of each of the classifier and bring upon the result and observe their accuracies by assigning different weights to them. The best accuracy that was achieved by performing the above is = 83.205%. The paper also harnessed some tools to build graphs and word cloud based on the results.

B. Background and Previous Work

In a [1], João et Al's main objective is to enable medical practitioners to make prompt, well-informed decisions by identifying patterns that are hidden within the textual data. But the results highlight how difficult it is to obtain accurate diagnoses based just on keywords supplied by a Portuguese medical unit; the "K-Nearest Neighbor" algorithm is the only one that shows some effectiveness. In [2] James Mullenbach et Al's attempted to extract International Classification of Diseases (ICD) codes from clinical notes—a critical task for effective administration of healthcare.A brand-new method was put out that uses an attentional convolutional network

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:37:26 UTC from IEEE Xplore. Restrictions apply.

(CAML) to predict medical codes straight from clinical literature having a precision of 0.71 and a Micro-F1 score of 0.54 which surpasses previous works.

Ali et al [11] discusses supervised machine learning models that use high-dimensional omics data to predict drug response. The researchers also compared different machine learning models and omics measurements. The researchers find that feature selection techniques are necessary to identify the most predictive biomarkers. These researchers conclude that with more data machine learning models can be useful for drug response prediction. [12] explored the potential of NLP for automating the highlighting of new information in clinical notes, potentially improving healthcare efficiency and clinician decision-making using crucial updates in medical records. The authors argue that current methods are time-consuming and require significant human expertise.

In [3] Benicio et Al's offered a novel method for standardizing and retrieving patient data utilizing advanced Text Mining and Natural Language Processing (NLP) techniques from electronic patient records (EPRs). The research proves the usefulness of the strategy through thorough examination, including statistical comparisons between manually retrieved data and that processed using the proposed automated methodology. In [4] Shen F, Liu et Al's purpose were to create a common evaluation methodology for systems that extract FH data from clinical narratives. The task is divided into two smaller tasks: assigning living status, family side, and clinical observations to family members; and recognizing family member entities and clinical observations (diseases).The findings indicate that entity identification is not as challenging as relation extraction from FH.

Sezgin et al [13] explored the potential and feasibility of using NLP to extract medication and symptom information from unstructured patient data collected via a voice-interactive app. The analysis involved notes from 24 parents of children with special needs. The NLP pipeline used a zero-shot approach and achieved promising results (precision ¿ 0.65, recall ¿ 0.77, F1 ¿ 0.72) for medication and symptom extraction. Alfattni et al [14] evaluated the effectiveness of NLP and deep learning for extracting and linking drug information from clinical notes. The authors use 2018 n2c2 shared task data to propose an NLP system ("DrugEx") that utilizes various embedding techniques for drug entity recognition and relation extraction. Their findings suggest that deep learning with pre-trained word and character embeddings performs well, with additional insights on limitations of rule-based vs. context-aware LSTM methods for relation extraction.

In [5] Tal Baumel et Al's explored the complex process of assigning numerous diagnosis codes to patient records, which was difficult due to large number of codes.Utilizing the MIMIC II and III clinical datasets, the paper demonstrates HA-GRU's superior performance against SVM, CBOW, and CNN classification techniques, especially in identifying rolled-up ICD9 codes. In [6] Y. Wang et Al's goals were to investigate current studies on clinical IE applications. 263 pertinent publications were found after a comprehensive search across many databases was conducted between January 2009 and September 2016. This review provides information on this gap and suggests some ways to close it.

Rybinski et al [15] explored automated methods for extracting family history information from clinical notes using NLP. The authors present a system that combines transformers for disease mention detection and rule-based methods for family member identification. Their approach achieved an F1 score of 81.63% on a public dataset. Gholipour et al [16] investigated the use of NLP for automatically extracting cancer-related information from clinical notes. By analyzing previous research, authors found that rule-based NLP algorithms were most frequent, with a focus on accuracy and sensitivity. These researchers observed a preference for SNOMED-CT and UMLS terminologies for identifying cancer concepts. The authors suggest applying these methods to other diseases as well.

In [7] D. Antons et Al's describes the procedure, which uses techniques like tokenization and tf-idf measurements for content analysis and data preprocessing. Dictionary-based and algorithmic text mining techniques are distinguished, with examples including clustering and classification provided. It draws attention to the development of text mining algorithms and results as well as issues with reporting transparency. In [8] M. Oghbaie et Al highlighted how crucial information retrieval (IR) is to managing large amounts of data, as information extraction and pattern identification are aided by document classification and clustering.Vector Space Model (VSM) is presented, and other term weighting techniques such as tf-idf are addressed. The study suggests a novel similarity metric that improves text performance by taking into account both term weights and the total number of terms in documents.

Harrison et al [17] explored the use of NLP techniques for analyzing large text datasets using open-source software and freely available data of patient reviews for various drugs.While sentiment analysis skews towards particular drugs, unsupervised learning (LDA) identifies clusters of drugs based on review characteristics. Supervised machine learning models achieved moderate accuracy in predicting drug review ratings. Kersloot et al [18] analyzes 77 studies from a pool of 2355 and found a lack of consistent reporting, with over a quarter not evaluating their algorithms. Most studies relied on internal data and manual annotations. The authors propose 16 recommendations to improve future research, emphasizing standardized reporting, data use, evaluation, and generalizability.

In [9] N.Bhat et Al focuses on strategies for producing captions that can combine novel approaches to visual

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:37:26 UTC from IEEE Xplore. Restrictions apply.

attention and maximize current notions. The survey highlights computationally feasible and optimal techniques that can be supported on various platforms, including lightweight devices like smartphones, and emphasizes the significance of adaptability and effectiveness of existing research in real-life applications.In [10] E.Mahajan et Al's used Convolutional Neural Network (CNN), Bi-directional Long Short-Term Memory (BiLSTM), Bi-directional Gated Recurrent Unit (Bi-GRU), and Long Short-Term Memory (LSTM) approaches in their suggested model to improve overall performance. %.

Abhishek malik al [19] evaluated the performance of the suggested method against a number of state-of-the-art models that have recently been proposed, deep learning models, pre-trained Bidirectional Encoder Representations from Transformers (BERT) using transfer learning, and several standard Machine Learning baseline models. Four datasets from various areas are used to evaluate these models for both training and testing.Sanjay K et al [20] were following processing, the features are used to produce balanced, well-labeled datasets that are categorized according to predetermined time intervals. These datasets are then used to train several machine learning classifiers, which are subsequently combined in a weighted way to create a better recommendation system based on fusion.

II. Evaluation Metrics

Since the paper has adapted several approaches to solve the problem, several metrics had to be figured out on the basis of which it could figure out the best approach and in which direction to proceed to improve the results. Hence The paper adopted the standard metrics using which AI models are evaluated. These include Accuracy, Hamming Loss, Precision, Recall and F1- score. These metrics provide an insight on the performance of each of the approach. Accuracy quantifies the overall correct predictions among all cases, offering a snapshot of model performance.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Accuracy} = 1 - \text{Hamming Loss}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN}$$

Fig. 1. Evaluation Metrics Formulae

Hamming Loss measures the fraction of wrong labels to total labels, indicating label-specific errors in multi-label settings. Precision assesses the accuracy of positive predictions, showing how many selected items are relevant. Recall, or Sensitivity, evaluates the model's ability to capture all relevant instances. F1-Score harmonizes precision and recall into one metric, providing a balanced view of a model's precision and recall capabilities, especially useful in imbalanced datasets. The formulas are given in the Fig. 1. The terms have the following meaning TP = True Positive TN = True Negative FP = False Positive FN = False Negative

III. Dataset

The paper used mtsamples.com, a well-known large database of medical report transcriptions. Both aspiring and experienced medical transcriptionists can use these examples for their everyday transcribe requirements.

The data was quiet unstructured and it had the following parameters (1) Description: This column provided a summary of the medical transcription content, offering insight into the nature of the medical case or scenario documented within the transcription. (2) Medical Speciality: Each transcription is classified under a specific medical specialty This classification helps in organizing and filtering transcriptions based on the specialty they belong to. (4) Transcription: This is the heart of the dataset, this column contains the actual medical transcriptions in text format.These transcriptions detail the clinical encounters, patient histories, examination findings,etc. (3) Keywords : Extracted from the transcriptions, this column includes relevant keywords or phrases that capture essential concepts, conditions, procedures, or findings documented within each transcription.

The problem with the above dataset was it was highly unstructured so healthcare repository dataset was chosen which also had the same parameters but it was more structured and so it was easier to work with it,

IV. Methodology

A. Pre-Processing

Tokenisation : Tokenization is the process of breaking down a piece of text into smaller units called tokens. These tokens can be words, phrases, symbols, or other meaningful elements that serve as the basic building blocks of the text. It plays a fundamental role in various natural language processing tasks, such as text classification, sentiment analysis, and machine translation, by enabling efficient manipulation and understanding of textual data.

Pad_sequence: The pad_sequences function is used for padding sequences to ensure that they all have the same length. In NLP, when working with text data, it is common to have variable-length sequences of words. However, neural networks typically require fixed-length input sequences. The pad_sequences function is used to add padding (zeros) to shorter sequences or truncate longer sequences to achieve a consistent length. TfidfTransformer: It is commonly used in text processing to convert raw token counts into a numerical feature representation that takes into account the importance of terms in a document corpus.The paper used

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:37:26 UTC from IEEE Xplore. Restrictions apply.

TABLE I
EVALUATION METRICS FORMULAE

Clinical Dataset Details	Training Size	Testing Size	Tag Size
MedicalRecords 21' - MT Samples	1980	594	3007
Healthcare Insight Repository	1504	451	3898

this for transforming because it is light weight unlike the BERT Model. CountVectorizer: CountVectorizer is a feature extraction method that converts a collection of text documents to a matrix of token counts. Each row in the matrix represents a document, and each column represents a unique word in the corpus. The values in the matrix are the counts of each word in each document. It is commonly used as a preprocessing step in text classification tasks.

Stopwords Removal : Stopwords is a set of English stop words (common words that are often considered non-informative). The function iterates through the words in the text, and if a word is in the set of stop words, it is removed from the list of words. BadSymbolsRemoval: Remove the bad symbols which have no meaning in the context. In this context special symbols have no meaning so remove it.
To implement the preprocessing techniques keras and sklearn(python libraries) were used.Keras offers tokeniser and pad]sequences for tokenisation and pad_sequences respectively.For TfidfTransformation and CountVectorisation feature extraction library from sklearn module were used.

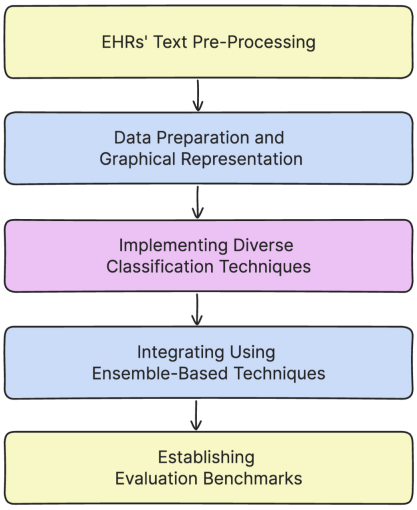


Fig. 2. Medical Term Extraction Visualization

B. Classifiers

1) One Vs Rest Classfier: The paper utilize the One-vs-Rest classifier in the study for its adept handling of multi-class classification challenges. Its strategic approach of breaking down intricate, multi-class tasks into several binary classification problems significantly streamlines both the training and prediction processes, making it an ideal choice for the research on extracting medically relevant terms from datasets.

The One-vs-Rest classifier operates by constructing a distinct binary classifier for each class, determining the membership of a given sample by comparing its prediction scores across these classifiers to select the class with the highest probability. The reduced accuracy observed with the One-vs-Rest classifier in the context may stem from the inherent imbalance in the dataset, where certain classes are underrepresented. This imbalance can lead to biased predictions towards more dominant classes, diminishing the model's ability to accurately classify minority class instances.

2) SGD Classifier: Next Up the paper focuses on an SGD classifier on the data. SGD classifier employs a stochastic approach, updating parameters using either single training examples or small batches, making it computationally efficient and suitable for large datasets. SGD is efficient as it uses small batch updates over dataset . It has high scalability and can easily handle high dimensional data .
It also allows us to apply various loss functions and regularization which makes it flexible with data. It Had a better accuracy then 1 vs rest classifier on the mtsamples dataset (70.43%). While applying SGD,be really careful with the input parameters as it is highly sensitive hence, careful pre-processing had to be done. Since it works in small batches so it is difficult for it to adapt to complex relationships. Also it is prone to over-fitting.

3) XGB Boost Classifier: The accuracy from SGD classifier was 70.43% so there was a huge room for improvement. So next up is to use XGB Boost classifier which is popular for its speed and performance in data tasks. It implements gradient boosting with enhancements like parallelized tree building and regularization techniques. The XGB classifier offers excellent in working on high dimensional data making it suitable for complex and diverse nature of clinical text. Also it has a very high ability to capture non linear relationships and interaction between features which allows for more accurate predictions and lastly it is scalable to large dataset. Got an accuracy of 73.2% which is more than both SGD and One Vs rest Classifiers.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:37:26 UTC from IEEE Xplore. Restrictions apply.

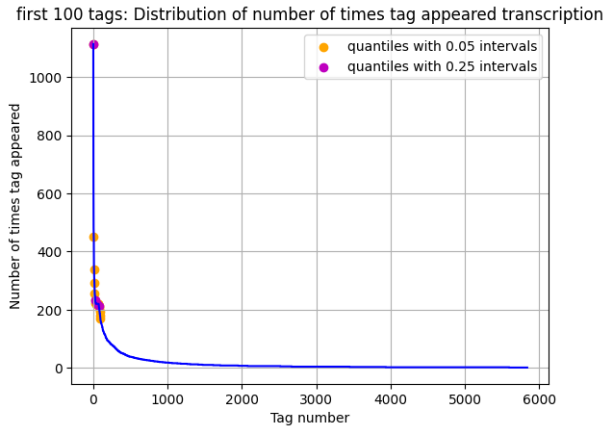


Fig. 3. Tag Frequency Distribution with Quantile Markers

Despite its strengths, XGBoost has its challenges when applied to clinical text data. One notable drawback is its difficulty in providing clear interpretations of its predictions, which can be crucial in complex medical scenarios where understanding the reasoning behind a prediction is essential. Additionally, fine-tuning XGBoost's parameters is a meticulous process, and improper tuning could result in models that either overfit or underfit the data, particularly when dealing with the nuanced and varied nature of clinical text.

4) MLP Classifier: Even after applying XGB Boost classifier the accuracy was still 73%, hence an MLP classifier which is an artificial neural network architecture renowned for its versatility in various machine learning tasks, including text classification was incorported. Their structure mimics the human brain, consisting of an input layer, one or more hidden layers with interconnected processing units (neurons), and an output layer. The major advantage of using MLP is its ability to model non-linear relationships between features and the target variables. This allows them to capture intricate patterns in data that linear models might miss.

However due to it's complex structure, it can lead to over fitting, especially with limited training data. Furthermore training MLPs often requires careful hyper-parameter tuning, which can be time-consuming and computationally expensive. Despite that, on the mtsamples dataset, MLP gives an accuracy of 82.69% which is way higher than other classifiers like SGD (70.43%), One Vs Rest (64.23%) and XGB (73.46%).

First the performance of the simple lone algorithms was analysed by performing tests on the same set of data in all of them. These include MLP, one vs rest ,SGD, XGB Boost Classifiers. However their performance was not upto the mark. Hence a different approach was taken to combine multiple classifiers and build upon their strengths to enhance

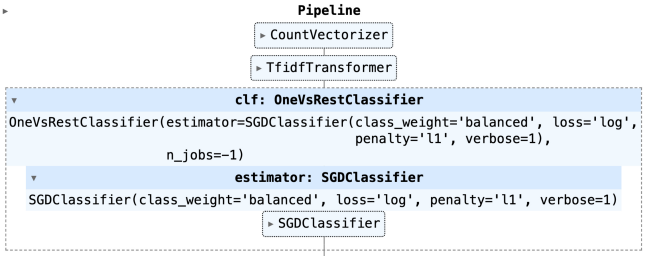


Fig. 4. Pipeline Visualization: Text Feature Extraction and Classification

the overall performance.

Various classification mechanisms can be utilised to enhance the performance of the model using the various python libraries like Sklearn and Keras. The models are inbuilt in into these libraries and the research paper has utilised these to come to results.

C. Ensemble Learning

Onto the final model, the paper used ensemble learning, it is a machine learning technique that combines the predictions from multiple base models to improve the overall performance and robustness of the final model. It operates under the principle that combining diverse perspectives from multiple models can lead to more accurate and reliable predictions. Types of Ensemble learning are as follows:

- Voting Classifier
- Stacking (Meta Learning)
- Bagging (Bootstrap Aggregating)
- Blending
- Boosting

The key benefit of using ensemble learning lies in its ability to reduce variance. By combining multiple models, it can average out the errors of individual models which can enhance robustness to overfitting. However, ensemble learning does come with certain drawbacks. The increased complexity of combining multiple models can lead to higher computational costs for training and prediction compared to simpler models. Furthermore, choosing the most appropriate base models and combining their predictions effectively requires careful consideration and experimentation.

The paper used Voting classifier for the model by using SGD, MLP and One vs Rest classifier as estimators. Some of the results obtained were: 78.2% (One vs Rest, SGD, MLP), 66.53% (One vs Rest, SGD), 82.17% (One vs Rest, MLP), 81.92% (MLP, XGB) and so on with weights equal to 1.

The paper also used different weights for different models to find the best model configurations. Out of all the tried configurations, found out that using MLP and SGD as estimators with weights 5 and 1 gives the best result of 83.21%.
The voting classifier has two types of voting's viz. Hard Voting and Soft voting. In Hard Voting what is done is predicting the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:37:26 UTC from IEEE Xplore. Restrictions apply.

TABLE II
CLASSIFIER DETAILS ALONG WITH EVALUATION METRICS

Classifier Details	Accuracy	Hamming Score	Precision	Recall	F1- Score
One vs Rest Classifier	64.23%	35.77%	0.5302	0.5801	0.5453
SGD Classifier	70.43%	29.57%	0.7043	0.7043	0.7043
XGB Boost Classifier	73.46%	26.53%	0.7346	0.7346	0.7346
MLP Classifier	82.69%	17.30%	0.7051	0.8269	0.7196
VEC (One vs Rest, SGD, MLP)	78.20%	21.80%	0.7820	0.7121	0.7820
VEC (One vs Rest, SGD)	66.53%	33.47%	0.6653	0.6653	0.6653
VEC (One vs Rest, MLP)	82.17%	17.83%	0.8217	0.7802	0.8217
VEC (MLP, XGB)	81.92%	18.08%	0.8192	0.7366	0.8192
VEC (MLP, SGD)	81.15%	18.85%	0.8110	0.8008	0.8110
VEC (MLP = 3, SGD = 1)	82.69%	17.30%	0.7014	0.8269	0.7196
VEC (MLP = 5, SGD = 1)	83.21%	16.79%	0.8320	0.8224	0.8121

data into classification tasks. The final predicted value is taken as the majority votes among the classifiers. Hence the class with the maximum number of votes is the final output. In Soft voting, every base model provides a probability distribution over the possible outcomes rather than a single prediction. The final output is determined by averaging the predicted probabilities from all individual models. The class having the highest average probability across all the models is selected as the final prediction.

V. FINDINGS

After applying different models to the clinical dataset, the best results were obtained while working with a voting ensemble classifier which combines MLP Neural Networks and SGD classifier with weights set to 5 and 1 respectively. The maximum accuracy was 83.25%.

The research findings offer a detailed analysis of machine learning classifiers and ensemble methods applied to medical text classification tasks. It emphasises the importance of preprocessing techniques such as tokenisation, sequence padding, and feature extraction using TF-IDF and CountVectorizer to effectively prepare text data for analysis. Among the classifiers tested—including One-vs-Rest, SGD, XGBoost, and MLP—the MLP classifier emerged as the top performer with an accuracy of 82.69%, leveraging its ability to capture complex data patterns.

The study also explored ensemble learning techniques, demonstrating the effectiveness of a Voting ensemble approach that strategically combined MLP and SGD models to achieve an impressive accuracy of 83.21%. It also mentions the challenges associated with model interpretability and scalability, particularly with classifiers like XGBoost, underscoring the importance of balanced model selection for real-world medical text analysis. These insights provide valuable considerations for deploying machine learning solutions in healthcare applications.

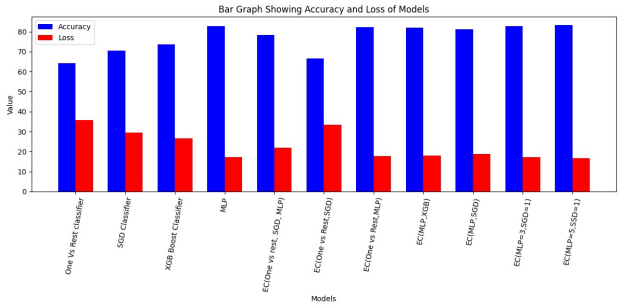


Fig. 5. Performance Visualizer

VI. FUTURE WORK

The paper have successfully extracted the keywords from the clinical text, and Hence the symptoms can be extracted from the clinical text. Thus, using these symptoms the model should be able to figure out the disease with a viable accuracy. Further corresponding to the symptoms the the model will recommend medicines/prescriptions. Future work also includes development of an app that will store the prescriptions and medicines taken by the user.

This will require a local database that will be useful for future purposes for the doctor. Also, developing frameworks that ensure seamless integration with diverse electronic health record systems will be a potential useful feature as well.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:37:26 UTC from IEEE Xplore. Restrictions apply.

Hyperparameter tuning for each individual classifier used as well as the ensemble classifier to boost the accuracy of the final model.Experimentation with different ensemble techniques such as stacking,blending or boosting can also give interesting results regarding the perfromance of the final model.

ACKNOWLEDGMENT

To express gratitude, the authors acknowledge the valuable tools and techniques that supported the research. The paper employed various machine learning tools including Stochastic Gradient Descent, One-vs-Rest, XGBoost, Multi-Layer Perceptron, and ensemble learning, which were trained on mtsamples dataset from kaggle. These diverse approaches enabled a comprehensive exploration of modeling techniques which lead to the successful extraction of relevant keywords with an accuracy of 83.205%.

REFERENCES

[1] João Ribeiro, Júlio Duarte, Filipe Portela, Manuel F. Santos. Automatically detect diagnostic patterns based on clinical notes through Text Mining. The 3rd International Workshop on Healthcare Interoperability and Pervasive Intelligent Systems (HiPIS 2019) November 4-7, 2019, Coimbra, Portugal Automatically detect diagnostic patterns based.

[2] James Mullenbach, Sarah Wiegreffe, Jon Duke, Jimeng Sun, and Jacob Eisenstein. 2018. Explainable Prediction of Medical Codes from Clinical Text. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 1101–1111, New Orleans, Louisiana. Association for Computational Linguistics.

[3] Benicio, Diego Henrique Pegado and Xavier-Júnior, João Calos and Paiva, Kairon Ramon Sabino de and Camargo, Juliana Dantas De Araújo Sant, Applying Text Mining and Natural Language Processing to Electronic Medical Records for Extracting and Transforming Texts into Structured Data. Available at SSRN: https://ssrn.com/abstract=3991515 or http://dx.doi.org/10.2139/ssrn.3991515

[4] Shen F, Liu S, Fu S, Wang Y, Henry S, Uzuner O, et al. Family history extraction from synthetic clinical narratives using natural language processing: overview and evaluation of a challenge data set and solutions for the 2019 National NLP.Clinical Challenges (n2c2)/Open Health Natural Language Processing (OHNLP) competition. JMIR Med Inform 2021 Jan 27;9(1): e24008 [FREE Full text] [doi: 10.2196/24008] [Medline: 33502329]

[5] Tal Baumel, Jumana Nassour-Kassis, Michael Elhadad, and Noémie Elhadad. 2018. Multi-label classification of patient notes a case study on ICD code assignment. In AAAI Workshop on Health Intelligence.

[6] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, H. Liu, Clinical information extraction applications: A literature review, Journal of Biomedical Informatics, Vol. 77 (2018) pp. 34–49.

[7] D. Antons, E. Gr¨unwald, P. Cichy, T. O. Salge, The application of text 385 mining methods in innovation research: current state, evolution patterns, and development priorities, R&D Management Vol. 50 (3) (2020) pp. 329– 351

[8] M. Oghbaie, M. Mohammadi Zanjireh, Pairwise document similarity measure based on present term set, Journal Big Data Vol. 5 (52) (2018) pp.

[9] A. Bhowmik, S. Kumar, N. Bhat, "Evolution of automatic visual description techniques-a methodological survey". Multimedia Tools and Applications, 80(18), 28015-28059 (2021)

[10] E. Mahajan, H. Mahajan, S. Kumar, "EnsMulHateCyb: Multilingual hate speech and cyberbully detection in online social media", Expert Systems with Applications 236, 121228, (2024).

[11] Ali M, Aittokallio T. Machine learning and feature selection for drug response prediction in precision oncology applications. Biophys Rev. 2019 Feb;11(1):31-39. doi: 10.1007/s12551-018-0446-z. Epub 2018 Aug 10. PMID: 30097794; PMCID: PMC6381361.

[12] S. Dhar, S. Das and S. Majumder, "A Study on NLP Based Approach in AI and Text Data Mining for Automated Highlighting of New Information in Clinical Notes," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 966-972, doi: 10.1109/ICICCS53718.2022.9788312.

[13] Sezgin E, Hussain SA, Rust S, Huang Y. Extracting Medical Information From Free-Text and Unstructured Patient-Generated Health Data Using Natural Language Processing Methods: Feasibility Study With Real-world Data. JMIR Form Res. 2023 Mar 7;7:e43014. doi: 10.2196/43014. PMID: 36881467; PMCID: PMC10031450.

[14] Alfattni G, Belousov M, Peek N, Nenadic G. Extracting Drug Names and Associated Attributes From Discharge Summaries: Text Mining Study. JMIR Med Inform. 2021 May 5;9(5):e24678. doi: 10.2196/24678. PMID: 33949962; PMCID: PMC8135022.

[15] Rybinski M, Dai X, Singh S, Karimi S, Nguyen A. Extracting Family History Information From Electronic Health Records: Natural Language Processing Analysis. JMIR Med Inform. 2021 Apr 30;9(4):e24020. doi: 10.2196/24020. Erratum in: JMIR Med Inform. 2021 May 3;9(5):e30153. PMID: 33664015; PMCID: PMC8092929.

[16] Gholipour, M., Khajouei, R., Amiri, P. et al. Extracting cancer concepts from clinical notes using natural language processing: a systematic review. BMC Bioinformatics 24, 405 (2023). https://doi.org/10.1186/s12859-023-05480-0

[17] Harrison, C.J., Sidey-Gibbons, C.J. Machine learning in medicine: a practical introduction to natural language processing. BMC Med Res Methodol 21, 158 (2021). https://doi.org/10.1186/s12874-021-01347-1

[18] Kersloot, M.G., van Putten, F.J.P., Abu-Hanna, A. et al. Natural language processing algorithms for mapping clinical text fragments onto ontology concepts: a systematic review and recommendations for future studies. J Biomed Semant 11, 14 (2020).

[19] A. Mallik, S. Kumar, "Word2Vec and LSTM based deep learning technique for context-free fake news detection". Multimedia Tools and Applications, 83, 919–940, (2024)

[20] A. Kumar, D.K. Jain, A. Mallik, S. Kumar, "Modified node2vec and attention based fusion framework for next POI recommendation". Information Fusion, 101, 101998 (2024)

LIST OF ABBREVIATIONS

- NLP : Natural Language Processing
- EHRs : Electronic Health Records
- EPRs: Electronic Patient Records
- SGD: Stochastic Gradient Descent
- MLP: Multi Layer Perceptron
- TF-IDF: Term Frequency Inverse Document Frequency
- XGB: Extreme Gradient Boosting
- VEC: Voting Ensemble Classifier
- ICD: International Classification of Diseases
- CAML: Categorical Abstract Machine Language
- UML: Unified Modeling Language
- SNOMED-CT: Systematized Medical Nomenclature for Medicine–Clinical Terminology
- VSM: Vector Space Model
- SVM: Support Vector Machine
- CNN: Convolutional Neural Network

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 08,2024 at 09:37:26 UTC from IEEE Xplore. Restrictions apply.

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-19646-2

VEDANet: A dense blocked network for visual emotion analysis in multimedia retrieval

Kirti Sharma[1] · Rainu Nandal[1] · Shailender Kumar[2] · Kamaldeep Joshi[1]

Received: 10 August 2023 / Revised: 10 March 2024 / Accepted: 7 June 2024
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract
This research focuses on addressing the challenges of visual emotion analysis and emphasizes the need for improved comprehension and categorization of emotions by machines. The objective is to develop an efficient architecture for emotion classification in real-world scenarios encountered in multimedia retrieval tasks, considering factors like illumination, occlusion, pose variations, small face sizes, multimodal detection, and big data issues. The proposed Dense Blocked Network-based VEDANet architecture surpasses the state-of-the-art on benchmark datasets by leveraging pre-trained CNN architectures to recognize facial features and extract metadata. The exploration of 128 descriptors from a deep residual network further enhances the operations. Using the VEDANet framework, emotions are classified into seven categories with superior accuracy compared to conventional approaches. The study investigates the efficacy of an over-the-top optimization (OTO) layer to enhance emotion classification. The proposed model achieves impressive accuracy scores on diverse datasets, including AffectNet, Google FEC, Yale Face DB, and FER2013, with percentages of 87.30%, 92.75%, 95.07%, and 90.53% respectively. Real-time performance is achieved with exceptional accuracy of 93.46% on live frames, while minimizing turn-around time by optimizing network size and parameters. This research contributes to the advancement of visual emotion analysis, providing a comprehensive and efficient solution for AI-based emotion detection in various applications.

Keywords Sentiment Analysis · Emotion Detection · Facial Emotion Recognition · FER · Deep Learning Architectures

✉ Rainu Nandal
rainunandal11@gmail.com

Kirti Sharma
krtbhardwaj1@gmail.com

Shailender Kumar
shailenderkumar@dce.ac.in

Kamaldeep Joshi
kamalmintwal@gmail.com

[1] CSE Department, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, Haryana, India

[2] CSE Department, Delhi Technological University, New Delhi, India

1 Introduction

Identification of a person's emotional state is crucial to several study fields, including political science, computer vision, psychology, and the sciences of medicine. Information reception, attitude development, and decision-making are all influenced by emotions. Sentiments exist on a deeper level of the human mind, whereas emotions at a given instant come from the subject's feeling for an item or person. Sentiments form part of a person's character and stay with them for a long time, whereas emotions are exhibited for a shorter time (typically instantaneously) as a result of an action or occurrence that triggers an emotional outburst.

Cognitive human emotion recognition and machine intelligence (AI), often known as human-computer interaction, are two subfields of emotion detection in academia. Facial landmarks, body language [1], and voice can reflect human emotions [2]. Mehrabian's 1967 research found 55% of emotional input was visual, 38% was aural, and 7% was spoken [3]. These findings still hold. Most research focus on facial expressions because they are the initial sign of emotional state transfer. The face displays emotions and intentions in social interactions and conflicts, which others see.

Advanced robotic and operational solutions will be based on face expressions. Human-based information extraction processed and classified into logical units will power robotic medical operations and sophisticated autonomous driving assistance technologies. Extracting features from one face compared to another to better categorisation is difficult and sensitive.

Ekman and Freisen were among the first scientists to examine face expression, establishing FACS (Facial Action Coding System) in 1978 to classify facial motions by AUs (Action Units). One or more face muscles are combined into 46 Action Units [4]. Geometric characteristics like Gabor wavelets [6], local binary patterns [5], and facial action units (FACs) [4] are already used to extract face features. However, categorising and identifying human emotions is difficult for a machine. Automatic Facial Emotion Detection is the most studied modality [7], but it's hard since everyone exhibits emotion differently. Researchers should consider the real-world challenges head position, brightness, backdrop, and occlusion. The current state-of-the-art, face landmark model [46] and face augmentation [47] approaches, addresses some of these difficulties but needs to be improved.

Corporate and academic interest in Deep Neural Networks (DNN) shows its efficacy. To comprehend human emotions, researchers are leveraging architectures such the deep convolutional neural network (DCNN), recurrent neural network (RNN), and attention networks, which enable automatic feature extraction and categorization [46]. CNN's key benefit over its forerunners is that it can recognise essential traits automatically, without human intervention and give better performance comparatively. Researchers have put forth a lot of effort to develop deep network architectures, going so far as to include blocks like the Squeeze-and Excitation Block to boost accuracy even more. Hu et al. (2018) used channel-wise attention with a squeeze and excitation network (SENet) across a deep residual network. The attention-module rearranges channels to teach the residual block quality features [8].

Liu et al. (2019) used Local Binary Pattern (LBP) and 2D-Gabor to extract feature discrimination using an ELM-SVM framework that achieved 85% accuracy on CK++ dataset [9]. Deeb et al. (2022) proposed the human emotion identification framework using HAAR, LDA, and IBH-based Extreme Learning Machine (ELM) classifier to detect the 7 fundamental emotions in real-time face emotion classification. The algorithm's accuracy on the

KDEF, CK+, and JAFFE (Japanese Female Facial Expression) public datasets reached 91.32, 97.62, and 97.78% respectively [10]. Appearance characteristics operate poorly with noisy facial images and are inconsistent with a number of facial changes (such as face size, head position, and face appearance) [40]. DenseNet methods achieved the highest results in terms of stability against noise perturbations in facial images [45]. Using findings from the most recent studies, we developed a framework that may be used in a variety of real-time applications in fields as diverse as teaching, autonomous vehicles, the entertainment industry, national security, and public safety.

In order to alleviate the burden of massive data training, it is prevalent to employ several well-known pre-trained CNN models, such as ResNet and VGGNet. So far, there has been no ground-breaking research into visual emotion identification that makes use of a pre-trained model to provide outcomes as similar to human emotions and thoughts as is practically possible. This is why in our system we employed ResNet as the underlying architecture for visual feature extraction. Features were retrieved by employing the transfer learning concept of computer vision by way of training deep CNNs on large data sets like ImageNet for VSA. During model training, there will be no need to start from scratch. We employed TL to minimise overfitting in our proposed study by bringing relevant information from one realm to another.

Previous research has focused on ideological input to the system. Our research provides an improved face value extraction approach based on integrating the most advanced pre-trained architectures in unity. Current researchers are attempting to achieve 100 percent efficiency; our study progresses towards this goal by delivering a more effective method to achieve a high rate of emotion recognition accuracy.

The primary contribution of the paper's proposed study is to provide a standard framework for detecting seven unique emotions—joyful, angry, disgusted, neutral, fearful, sad, and surprised—in face pictures and live broadcasts. The proposed work develops a pre-trained architecture for extracting visual face descriptors (FDs), which may be used for both face detection and recognition. Built on top of a modified Resnet (31layers) architecture, the model achieves good recognition accuracy owing to the inclusion of visual face descriptors (FD's), Bounding Boxes, and 68 Landmark points. The proposed work's main contribution is to explore the proposed framework on benchmark datasets and on real-time live frames, and present a novel VEDANet architecture (baseline: DenseNet-121) for Visual Emotion Detection and Analysis. The proposed architecture expanded by including an OTO Layer (Over the Top Optimization Layer) to boost the confidence score of the final result and speed up the precision with which emotions were identified.

In the final stage of the proposed work, the presented study contributed by providing the comparative analysis of the proposed model's (VEDANet) accuracy in emotion detection task with current SOTA and real-time live frames, demonstrating its competency to outperform. The investigation of proposed framework covered the interesting challenge of this field that is for balancing the accuracy over the emotion categories.

The paper is organized as follows. Section 1 serves as the introduction, providing an overview of the research and introducing the datasets used. In Section 2, we conduct a comprehensive review of existing approaches to emotion detection and analysis, focusing on their performance on standard benchmark datasets. Section 3 delves into the implementation process, including the data used, methodologies, and training hyperparameters. We also discuss the modifications made to enhance the performance of pre-existing models. The assessment metrics, results obtained from the experiments and their analysis are presented in Section 4. Finally, in Section 5, we draw conclusions based on the findings of the study.

1.1 Datasets used

1.1.1 Extended yale face database

This collection is made up of 16,128 facial pics that were all captured using the same light source. It has 28 unique subjects for 576 different viewing settings, including nine poses for each of the 64 different lighting conditions. Each facial image's actual dimensions are 320 x 243 pixels [11].

1.1.2 Google FEC database

700,000 triplets of distinct face crop photos made up the Google Facial Expression Comparison (FEC) [12] database. In each triplet, the most comparable pair of facial expressions are annotated [12] (Fig. 1).

1.1.3 Affectnet dataset

AffectNet [13], is actually the largest publicly accessible repository of face emotions which has ~ 440,000 in-the-wild face crop photos, each of which is manually tagged into one of eight categories of facial expression categories (Happy, Sad, Neutral, Surprise, Disgust, Fear, Contempt and Anger). In the presented work, to generalize the model seven emotion classes with annotations used except contempt emotion [13] (Fig. 2).

1.1.4 FER2013

The Face Emotion Recognition 2013 (FER-2013) dataset was developed by Pierre Luc Carrier and Aaron Courville for the facial expression recognition challenge of the ICML 2013 workshop [14]. The collection was created by crawling photographs that correspond to emotive keywords using the Google image search API [14] (Fig. 3).

1.1.5 Real-time wild dataset

We used a camera device (smartphones and webcam) to capture ongoing conversations and extracts ~350 frames from random instances. We exposed those frames to our proposed architecture to evaluate.



Fig. 1 Samples from google facial expression comparison dataset [12]



Fig. 2 Sample images from the AffectNet dataset (where 0,1,2,3,4,5,6,7 denoted emotion classes as neutral, happy, sad, surprise, fear, disgust, anger, contempt respectively) [13]



Fig. 3 Sample Images from FER2013 Dataset. (Having 35887 facial images= '8989 Happiness' + '6198 Neutral' + '6077 Sadness' + '5121 Fear' + '4953 Anger' + '4002 Surprise' + '547 Disgust') [14]

2 Background and related work

As per the recent past works in the field of emotion detection it was observed that CNNs are renowned for their ability to recognize and categorize images. A basic convolutional neural network typically comprises of neurons connected over several layers. These layers aggregate the incoming images and perform various processing layers on them.

Convolutional, pooling, and fully connected layers are the 3-types of layers that build up a basic CNN. The feature extraction process is carried out by the first two layers, which also introduce non-linearity and feature reduction to lessen overfitting. The last layer, referred to as a fully linked layer, aids in classifying data using attributes that were retrieved from earlier levels. The majority of the parameters are found in the fully linked (connected) layer. Certain architectures also aimed to reduce the number of parameters like in Inception V3 [15] the last layer of Global Mean (Average) Pooling operation is added. By averaging the feature map and turning it into a scalar vector, this layer shrinks the feature map. Modern CNN designs have demonstrated the utilization of residual modules [16] and depth wise-separable convolutions [17] to further minimize the parameters. The feature extraction task is split up and combined within the convolutional layer by the depth-wise-separable CNNs, which further reduces the parameters. A Mini-Xception CNN suggested by [18] that minimizes the number of parameters by using depth-wise separable dense blocks and eliminates fully connected layers. So, by taking motivation from [17, 18] research work, we additionally used depth-wise separable convolution layers rather than ordinary convolution layers that leads reduction in parameters.

Due to their built-in feature extraction method from images, Deep NNs (DNNs), mainly convolutional neural networks (CNNs), have gained attention in Facial Emotion Detection [19, 20]. In order to address the challenges with emotion detection, a few works have been reported with CNN [21–26]. The CNN's deeper model has been demonstrated to perform better at other image-processing tasks [27], however the Visual Emotion Detection methods now in use only take into account the first few layers of the CNN. The difficulties associated with the task of detecting emotions may be the facts underlying this. The first requirement for emotion recognition is a reasonably high-resolution image, which calls for working with high dimension data. Second, there are a tonne of hidden convolutional layers in a really deep CNN. It is difficult and unreliable to train the CNN with a large number of hidden layers. Additionally, the vanishing gradient issue means that just adding layers does not improve accuracy after a certain point [28]. The notion of skip-connection [29], which was first proposed in residual neural network (ResNet) [29], is one that stands out and is used by the majority of the later models. The fundamental concept behind a skip connection is to add a layer's input to the output after a few other layers. This gives the layer additional information and aids in resolving the vanishing gradient issue. There are already a few distinct ResNet models available, including ResNet-18 (17 convolutional layers), ResNet-34, ResNet-50, and ResNet-152.

According to DenseNet [30], which established extensive skip connections between layers, each layer now receives signals from its predecessors and uses their output to drive all succeeding layers. A single layer's input is blended with the channel concatenation of layers that came before it. DenseNet has [L (L + 1)/2] direct connections as opposed to the 'L' direct connections of conventional CNNs with 'L' layers. Because each layer has direct access to the layers above it, the network's information bottleneck is less severe. As a result, the model is substantially more compact and thinner, producing high computational efficiency. In addition to this, non-linearity on the kth layer is estimated using a non-linear function for this feature map and the concatenation of 0-to-(k -1) features. There are various variations of this model; the one we are interested in is the DenseNet-121, which uses activation functions and global level pooling to deploy 120- convolutional layers predominantly inside 4 blocks.

The deep CNN architectures have undergone a number of tweaks and training methods to improve accuracy [15, 31, 32]. VGG-16, Resnet-50, Resnet-152 [31], Inception-v3 [15], DenseNet121, and DenseNet-161 [32] are some of the often-used pre-trained DCNN models. But it takes a lot of data and powerful computing to train such a deep model.

So, we have picked some pitfalls from previous related work and tried to cover few challenges among them which are as follows: (i) Minimize error rate and increase accuracy for frontal face as well as side face images (ii) Use of Pre-trained architectures to minimize the pre-training computation time and cost. (iii) To provide a Real-time efficient and deployable Emotion Detection System. The proposed work discussed further in next section.

3 Proposed work strategy

The proposed stream-lined workflow represented in the block diagram as in Fig. 4. The pipelined steps involved in the workflow are as follows. (i) Input frame extracted from real-time device or external input source (datasets) (ii) Face Detection with pre-trained architectures (MobileNetV1- SSD, Tiny Yolo2) and Facial-Landmark model (with 68 view points) (ii) Visual Face Descriptors Extraction with VFDNet (Resnet- 31 layers) (iii) Visual

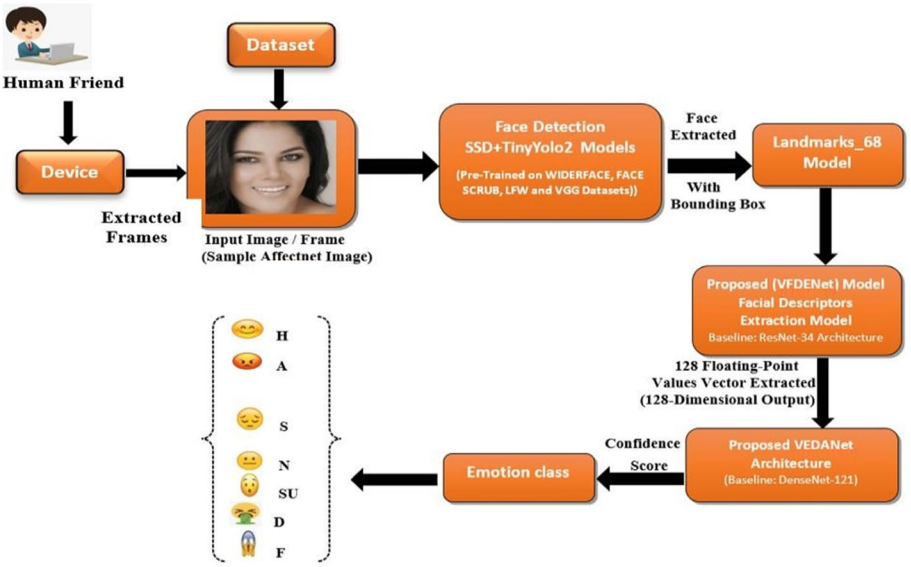


Fig. 4 Block diagram of proposed work for visual emotion detection and analysis

Emotion Detection and analysis with proposed VEDANet (baseline: Densenet-121) and extended the architecture by adding the OTO Layer in the end. The stepwise work progress is described below.

3.1 Face detection

MobileNetV1-SSD combines the depth-wise separable convolution from MobileNets with the Single Shot MultiBox Detector (SSD) framework, optimizing for speed and efficiency, ideal for mobile and real-time applications. Our face detection system utilizes the MobileNet SSD, a model optimized for efficient and speedy detection, integrated with the TensorFlow object detection API and refined using the extensive WIDERFACE dataset. This sophisticated network not only identifies the location of faces within an image by providing bounding box coordinates but also assesses the confidence level of each detection. Tiny Yolo2, a smaller and faster version of the YOLO (You Only Look Once) model, excels in real-time detection with reduced resource consumption. The synergy between these models arises from their complementary strengths: MobileNetV1-SSD's efficiency and accuracy in detecting larger faces, and Tiny Yolo2's capability to rapidly process images, albeit with slightly lower accuracy for small faces. This hybrid approach ensures robust face detection across various scenarios, balancing speed and precision.

Our face detection model has undergone training with a specialized dataset comprising approximately 14,000 images, each meticulously labeled with bounding boxes to ensure precise and accurate recognition capabilities. The enhanced model, trained to accurately predict bounding boxes that fully encapsulate facial feature points, delivers superior performance in landmark detection compared to using SSD MobileNet V1 alone. Leveraging the strengths of both the tiny face detector and SSD MobileNet, and augmented with a sophisticated landmark detection model, this approach ensures exhaustive facial feature coverage.

Such a comprehensive integration markedly boosts the system's efficiency under diverse conditions, guaranteeing precise and dependable face detection in various environments.

3.2 Visual facial features extraction

This stage of work implements a 68-point facial landmark detector that is lightweight, quick, and accurate. The standard pretrained model (face landmark-68 model) is just 351kb, while the tiny model (face landmark-68 tiny model) is only ~82kb. Both models make use of densely linked blocks and depth-wise separable convolutions. The models have been trained on a dataset of ~35k face images labelled with 68 facial-landmark points.

Deep convolutional neural networks have made a series of advancements in unconstrained face identification in current history, allowing us to deal with facial images with severe postures, lighting, and resolution changes. In the direction of further advancements to tackle these challenges, we have also designed VFDENet with 31 convolutional layers by considering ResNet as the baseline network shown in Fig. 5. It's a variant of the ResNet-34 network presented in Fig. 5, with three layers eliminated and the number of filters per layer halved to enhance operational efficiencies. This architecture is implemented to compute a face descriptor (a feature vector with 128 values) from any given face image, which is used to describe the characteristics of a person's face. The VFDENet architecture is shown in Fig. 5.

3.3 Visual emotion detection and analysis with proposed VEDANet architecture

After Face Detection and Facial Features Extraction (Landmark points and descriptors), this stage will evaluate the Emotion Classification based on proposed model VEDANet architecture (Table 1). Our proposed architecture exposed the exclusive outcomes comparative to SOTA due to the densely connected blocked network. The flow of working model is shown in Figs. 4 and 6. The Emotion Detection (ED) algorithm and Emotion Analysis (EA) algorithm is shown in Fig. 7.

Consider an input image x0 (128*128) that has been processed via the convolution layer. Proposed Architecture anatomy shown in Table 2, that takes the input image and turned half the size and doubled the filters and reduced number of parameters. The first convo block has 64 filters of 8*8 size and 2 strides. *Convo-layer is made up of feature maps and filters. Where feature maps are the outcome as per the weight applied to the filter and filters hold the input weights in accordance with that output value.* As in our architecture, 1st convo-layer is 8*8 and every dense block consists of two- convolutions with 1*1 and 3*3 sized kernels.

The Convolution layers are followed by 3*3 Max Pooling and 2 strides. The lth layer has $k0+k \times (l-1)$ input feature-maps, where k0 is the number of channels in the input

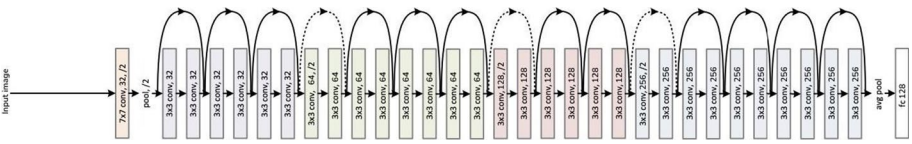


Fig. 5 Face descriptor extraction using the ResNet-31 layers' VFDENet architecture. The output is a vector of 128 floating-point values, and the dotted shortcuts add more dimensions

Table 1 VEDANet architecture: K = 32 represents the growth rate across the network

Layers	Output_Size	VEDANet Architecture (K=32)
Convolution	128*128	8 × 8 Convo_Layer (CL), Strides-2
Pooling	64*64	3 × 3 Max-Pool, Strides-2
Dense-Block (DB1)	64*64	$\begin{bmatrix} 1 * 1\text{conv} \\ 3 * 3\text{conv} \end{bmatrix} \times 6$
Transition Layer (1)	64*64	1×1 Convo_Layer (CL)
	32*32	2 × 2 Average-Pool, Strides-2
Dense-Block (DB2)	32*32	$\begin{bmatrix} 1 * 1\text{conv} \\ 3 * 3\text{conv} \end{bmatrix} \times 12$
Transition Layer (2)	32*32	1×1 Convo_Layer (CL)
	16*16	2 × 2 Average-Pool, Strides-2
Dense-Block (DB3)	16*16	$\begin{bmatrix} 1 * 1\text{conv} \\ 3 * 3\text{conv} \end{bmatrix} \times 24$
Transition Layer (3)	16*16	1×1 Convo_Layer (CL)
	8*8	2 × 2 Average-Pool, Strides-2
Dense-Block (DB4)	8*8	$\begin{bmatrix} 1 * 1\text{conv} \\ 3- * 3\text{conv} \end{bmatrix} \times 16$
Classification_Layer	1*1	Concatenation of 8*8 global average pooling & global max pooling
		FC Layer, Softmax



Fig. 6 (Over-the-top optimization) OTO layer workflow to ensure the highest confidence score and increase the success rate

For iteration X in (input) do	***OTO algorithm-***
X(static) >> SSD batch >> Xss	For iteration X in (input) processX(us) >> POP
(X << Xss)	batch >> XX
X(real frame) >> Yolo batch >> Xyo	***(XX-> emotion and confidence pair array)***
X<< Xyo	X(ms) >> POP batch >> XX
X(Xss, Xyo)	XX << POP batch
Input batch classifier	XX >>(Eus, Ems)NOW IF
Fbo << X (Xss, Xyo)	E1 << Eus
Face bounds Lx << fo(Lmo). 68 landmarks	E2<<EmsE1==E2
Dface << fo (Rx)	Do
Ex << Dense(Dface(Rx), X).	Conf<<ConfE1.ConfE2/attention-weighted averageT=0.6
Emotion class	Conf>T
E << arr(Ex)	Success
End	

Fig. 7 Emotion detection (ED) algorithm and emotion analysis (EA) algorithm. In this algorithm, the keywords are described as follows: Xss (output from SSD for input), Xyo (output from YOLO), Lx is landmarks of input x, fo (Lmo) is function of landmark modal, Dface is Face Descriptors, Lx landmarks of x. X(us) is Output of our modal, X(ms) is output of Microsoft Cognitive Services (MCS), POP is post processing output of each modal. E1 is emotion of our modal, E2 is emotion of MCS, T is threshold, Conf is confidence score

layer. The convo layer and pooling layer together yield k feature-maps. k equals 32 in the current network design. The network growth rate is referred to as the hyper-parameter k. *To reduce sample feature detection in feature maps, pooling is necessary. By summing up the existence of features in feature map patches, pooling layers offer a technique for down sampling feature maps.* To facilitate down-sampling in our architecture, a batch normalization layer and 1×1 convolutional layer, followed by a 2×2 average pooling layer used at each transition layer. In our implementation work, we considered that each 1×1 convolution produce ~4k feature-maps.

The pooling operation draws attention to the characteristics that overlap with filters by swiping a 2-D filter across the 3-D feature-map. The result of the pooling will be (***(h - f+ 1)/ s * (w - f + 1) * c***), if a feature map with dimensions of w* h* c is provided. Here, the feature map's height and width are denoted by h and w, respectively. The channels shown in the feature map, the filter's size and stride length denoted by c, f and s respectively. Additionally, the convolutional layer's most noticeable feature is used by the layer in max-pooling to operate on the feature map. The sharpest features on the image are the best lower-level features of the image, which can be extracted with the use of max-pooling. In proposed architecture, before the first dense block operations, a 3*3 max pooling operation is deployed to extract the most prominent feature from the feature map.

When using average pooling, the layer chooses the mean values of the items present in the feature map patch. The entire feature map is basically down-sampled to the average value recorded by the feature map's region. Each Transition Layer in the design uses 2*2 Average Pooling to assess the average value recorded by the region of the feature map. So, the most apparent feature of any patch is thus revealed by max-pooling, whereas average pooling reveals the average of the covered region. Max and Average Pooling Operations represented as below in Fig. 8.

Table 2 Comparative Study of performance evaluation of proposed VEDANet Architecture with State-of-the arts

Datasets	Models	Accuracy (%)
Affectnet	ER_MOPI (Used Meta-learning approach and included Occlusion, Pose and Illumination parameters) [34]	68%
	DenseSANet121 [35]	60.88
	Distilled CNN [37]	61.6%
	gCNN [39]	84.70%
	pACNN [39]	86.27%
	VEDANet (Ours)	**87.30%**
FER2013	CNN [33]	89.98%
	DeepEmotion [40]	70.02%
	Self-Attention ResNet (LHC-Net) [41]	75.42%
	VEDANet (Ours)	90.53%
Google FEC	DeepFEVER Distilled CNN [37]	86.5%
	Fine-tuned FaceNet [38]	81.8%
	VEDANet (Ours)	**92.75%**
Yale Face	HOG+VGG-Face feature-fusion model [36]	95.26%% (+/−2.5%) (5-Fold-Cross Validation)
	Mixed-feature model (LBP+2DPCA projection features + DDAG-based SVM) [42]	81.28% (4 classes: (happiness, neutral, sadness, surprise)
	VEDANet (Ours)	**95.07%**
Real-Time (Live Frames)	Multi Label Output [43]	79.26%
	Covariance Pooling [44]	87.0%
	VEDANet (Ours)	**93.46%**

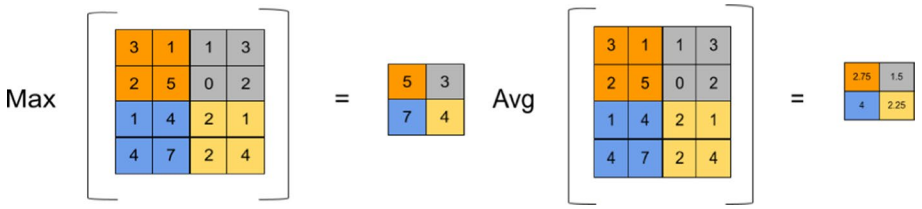


Fig. 8 Max pooling and average pooling operations

After receiving the input, each of the convo-block follow the further sequence as: Batch Normalization is followed by the ReLU activation function and the underlying Convo-2D layer.

Rectified Linear Unit (ReLU) activation is widely used in deep learning applications, and we adopted it for our work because of its many benefits. For newbies, the activation function f(x) is simply defined as the maximum value of the input x, making ReLU a computing juggernaut. Second, because it prevents saturation for high input values, ReLU eliminates the vanishing gradient problem that can hinder a network's ability to learn and effectively transmit gradients over several layers. In contrast, activation functions like the sigmoid or tanh could have to deal with this problem [48].

Additionally, deep networks benefit from ReLU's promotion of sparse representations. Keeping all negative activations at zero allows for just a subset of neurons to be active at once, which can help bring down the representation's complexity and simplify the model. The "dying ReLU" problem highlights a limitation of ReLU in which neurons might become locked in a "dead" state and fail to provide any output. To fix this, we first randomly sampled the weights from a normal distribution with mean 0 and standard deviation $\sqrt{\frac{2}{n {in} + n {out}}}$, where $n {in}$ and $n {out}$ are the input and output units for the layer.

Our architecture has four dense blocks, each has 2- convolutions with 1*1 and 3*3 sized kernels. Then, input is connected with the output tensor. And both of these convolutions will repeat for all the blocks (block-1:6times, block-2: 12times, Block-3: 24times and Block-4: 16times). Now, since the increasing number of channels can cause possible overfitting of the feature-map, we employ *Transition Layer* to cut-down the number of channels to half of the current channels. Then, we deployed the layer with 1*1 convolution and 2*2 Average Pooling Layer along with 2-Strides. Each of dense blocks are stacked with Transition Layers to make the model equably compact.

Later, till block-4, the output is structurally compact enough and hence, it is directly passed to classification layer as it's represented in Table 2 (VEDANet Architecture). In proposed architecture, at *Classification Layer,* the global average pooling layer averages each feature map and global max-pooling-layer extract maximum from each feature map then send the resulting vector (concatenation of both) direct into the softmax layer, *avoiding the possibility of overfitting*. Final output Layer as: FC Layer provides the emotion classification with respect to the input image X0 (128*128).

To further enhance our workflow for real-time live streams, we applied a hybrid approach to tap in potential of parallel processing, wherein we will parallelly assess our findings by taking advantage of another state-of-the-art cognitive service solution like, Microsoft Cognitive Services. In OTO layer we process the same input with Microsoft Cognitive Emotion Services and perform an attention-weighted average of the confidence scores to evaluate final confidence score (FCS). Further, we observed that if we tune the

threshold hyperparameter to be equal to 0.6 (i.e. T=0.6) the workflow achieves higher accuracy while classifying sub level classification predictions as inconclusive in turn decreasing the error rate at greater extent (~4-5%).

Finally, if the output of attention-weighted average is greater than the threshold (FCS >= T), we declare success. To assess the viability of our proposed model we also compared our model with SOTA benchmark datasets (AffectNet, FER2013, Yale, Google FEC) and Real- time live streaming.

3.3.1 Hardware GPU

The primary GPU used was NVIDIA RTX 4090 GPU with NVIDIA Ada Lovelace architecture and 24 GB of GDDR6X memory and a 384-bit memory interface. This configuration supports the 8-bit Float data type and contains Tensor Memory Accelerator (TMA) helpful for 8-bit matrix multiplication saving overhead of read/write indices.

4 Performance evaluation and comparative study

To evaluate the performance of the proposed architecture, the preferred evaluation matrices are Accuracy, Precision and F1-Score. Proposed Model proved the accuracy 87.30% on AffectNet dataset, 92.75% on Google FEC, 95.07% on Extended Yale Dataset and 90.53% on FER2013 dataset and 93.46 % on Real-Time Dataset respectively. The Confusion Matrix with respect to each benchmark datasets shown in Fig. 9. To measure how well a classification model does its job, statisticians create a Confusion matrix of size N by N, where N is the total number of classes being considered. The matrix evaluates the accuracy of the computational model's predictions vs the actual target values. Models with high TP and TN rates and low FP and FN rates are preferred.

Confusion matrices are the preferred metric for evaluating models when working with unbalanced datasets. The number of accurate and erroneous predictions produced by a classifier are summarized in a table called a confusion matrix. One uses it to evaluate how well a categorization model is doing its job.

Accuracy, precision, recall, and F1-score are just few of the performance indicators that may be calculated with this tool. Widespread use of confusion matrices can be attributed to their superiority over classification accuracy as a measure of a model's efficacy. For instance, in classification accuracy, we don't know how many cases were incorrectly labelled. Table 2 also shows a comparative study of the proposed working-model with current best practices. In this section, Table 3 represents the evaluated results of individual dataset's emotion classes with respect to the evaluation parameters accuracy, precision and F1-Score values.

From the Table 3, we can see that all emotion classes performed exceptionally well. If we look at the outcome of Fear and disgust, the performance of the model being a bit lower than other emotion classes, somehow, is related to the per instance physical ambiguity between the emotion classes. In human world, it is highly confusing to differentiate between the two emotions if one doesn't know the context of the dialogue resulting in those emotions. Nevertheless, in computational analysis this ambiguity is a result of low availability of defining state of the art datasets to train and improvise the model which in turn could possibly validate the contextual dependency to assess and segment these two emotions effectively.

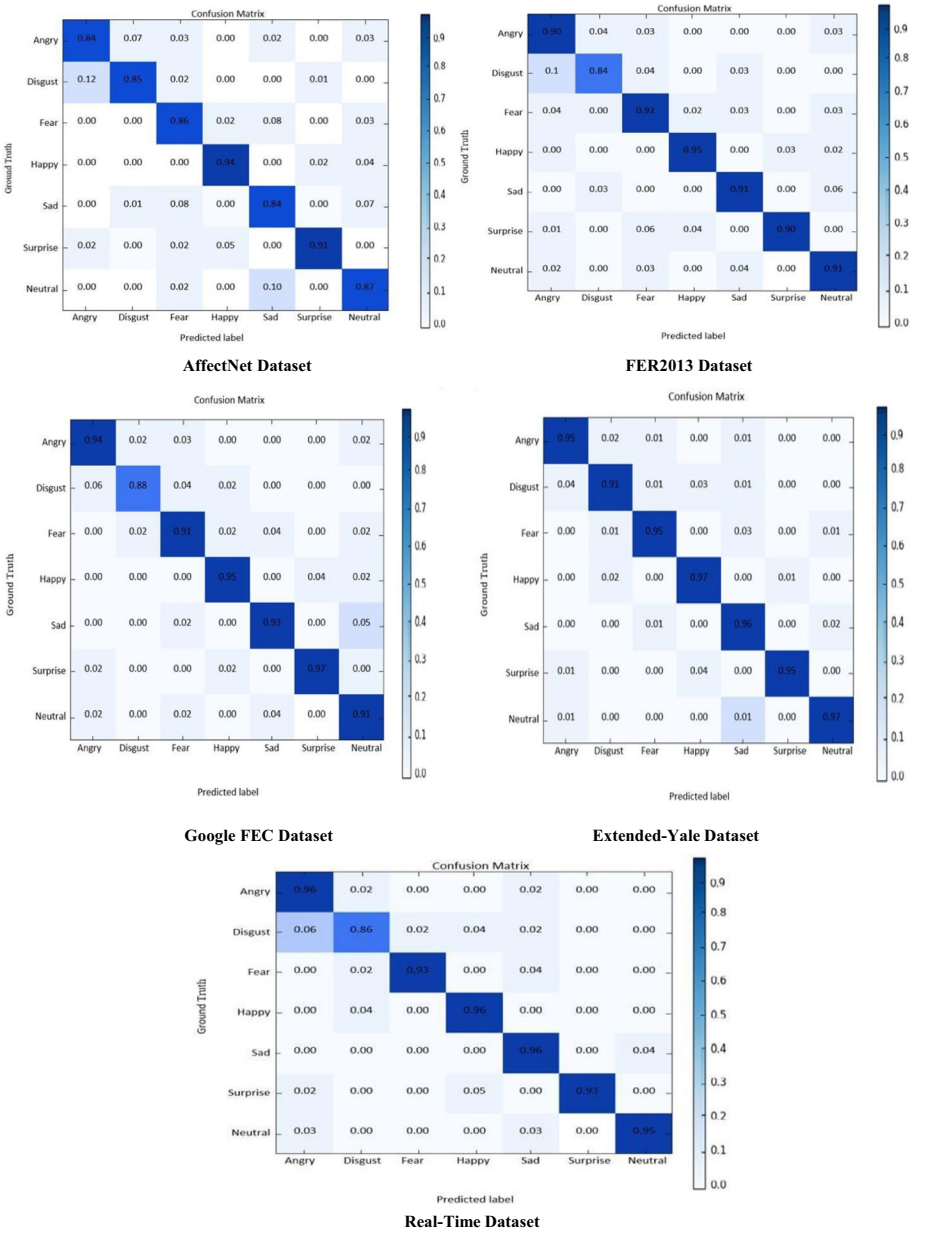


Fig. 9 Confusion matrices for affectnet, google FEC, extended yale, and real-time datasets

Accuracy is the classifier's prediction rate. The ratio of correct forecasts to total predictions. Unbalanced classes don't work with accuracy.

For unbalanced data, the model will predict that each point belongs to the majority class label, which is accurate but has drawbacks. The model is inaccurate. Achieving this level of accuracy is a hallmark of good prediction. For this reason, accuracy is

Table 3 Class wise (7 emotion classes) performance analysis of individual datasets (accuracy, precision and F1-score)

Emotion Classes (EC)		1	2	3	4	5	6	7	Overall
Datasets	Parameters	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral	Accuracy
	Accuracy	0.96	0.96	0.95	0.98	0.95	0.98	0.96	
AffectNet	Precision	0.85	0.92	0.84	0.93	0.81	0.96	0.82	87.30%
	Recall	0.92	0.91	0.93	0.96	0.92	0.94	0.94	
	F1-Score	0.85	0.88	0.85	0.93	0.83	0.93	0.85	
	Accuracy	0.97	0.97	0.97	0.98	0.97	0.98	0.97	
FER2013	Precision	0.89	0.91	0.86	0.94	0.92	0.97	0.86	
	Recall	0.95	0.91	0.96	0.98	0.94	0.94	0.95	90.53%
	F1-Score	0.9	0.87	0.89	0.95	0.91	0.93	0.88	
	Accuracy	0.98	0.98	0.97	0.98	0.98	0.99	0.97	
Google FEC	Precision	0.92	0.95	0.89	0.95	0.93	0.97	0.88	92.75%
	Recall	0.97	0.93	0.95	0.97	0.96	0.98	0.95	
	F1-Score	0.93	0.91	0.9	0.95	0.93	0.97	0.89	
	Accuracy	0.98	0.98	0.99	0.99	0.99	0.99	0.99	
Yale-Face	Precision	0.94	0.93	0.96	0.94	0.94	0.99	0.96	95.07%
	Recall	0.98	0.95	0.96	0.99	0.98	0.97	98	
	F1-Score	0.95	0.92	0.95	0.96	0.95	0.97	0.96	
	Accuracy	0.98	0.97	0.99	0.98	0.98	0.99	0.99	
Real- Time	Precision	0.91	0.92	0.98	0.92	0.9	1	0.95	93.46%
	Recall	0.97	0.92	0.95	0.98	0.98	0.96	0.97	
	F1-Score	0.93	0.89	0.95	0.94	0.93	0.96	0.95	

a suitable criterion for assessing classification problems that do not involve an imbalance of classes. For classification issues when the classes are evenly distributed, such as when no class imbalance exists, accuracy is a suitable measure of performance that can be leveraged to the proposed work. Accuracy can be evaluated as follows:

$$\boldsymbol{Accuracy} = \frac{\boldsymbol{TP + TN}}{\boldsymbol{TP + TN + FP + FN}} = \frac{\boldsymbol{CorrectPredictions}}{\boldsymbol{TotalPredictions}}$$

Accurate prediction may be quantified by its degree of precision. In layman's terms, it informs us what fraction of all optimistic forecasts come true. Precision should be exceptionally good (ideally 1). For situations where avoiding false positives is more important than avoiding false negatives, precision might be a helpful statistic. Classifying genuine emotions requires a high degree of precision. Precision can be evaluated by using below equation:

$$\boldsymbol{Precision} = \frac{\boldsymbol{TP}}{\boldsymbol{TP + FP}} = \frac{\boldsymbol{ActuallyPositvePredictions}}{\boldsymbol{TotalPositivePredictions}}$$

When both the number of false-positive and false-negative results matter, utilise the F1-Score. To evaluate Uneven Data, the F1-Score is more appropriate which can be evaluated as follows (Figs. 10 and 11):

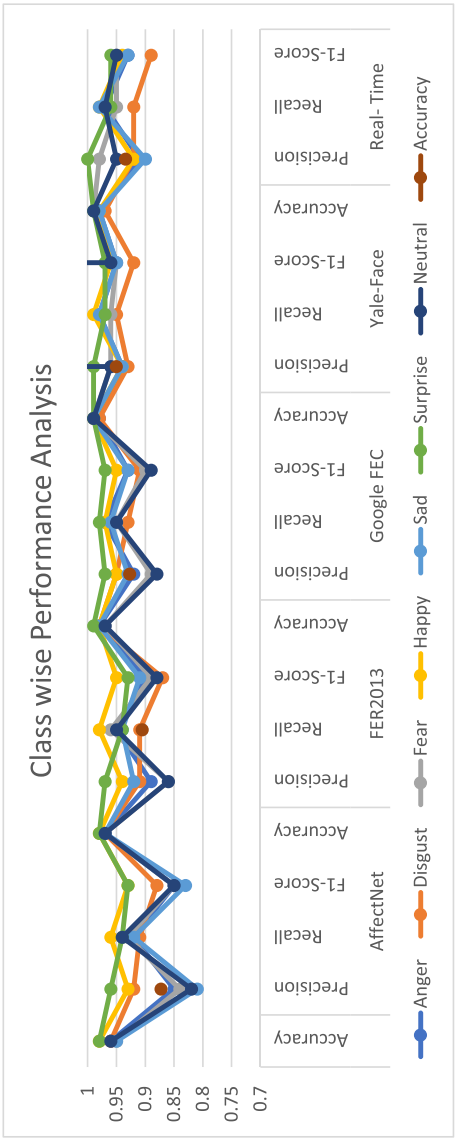


Fig. 10 Classwise emotion analysis of benchmark datasets

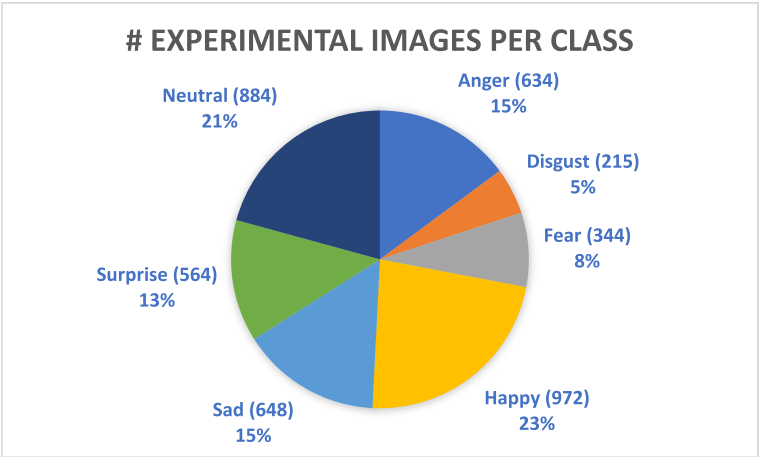


Fig. 11 Analysis of experimental images per emotion class

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5 Discussion

In this study, we introduced VEDANet, a novel deep learning architecture designed to significantly enhance visual emotion analysis. The core innovation of VEDANet lies in its One-To-Many (OTO) layer, a unique component that optimally maps single facial images to a spectrum of emotional states, acknowledging the complexity and subtlety of human emotions. This section delves into the methodology's effectiveness, the architectural novelties of VEDANet, its comparison with existing benchmarks, potential applications, limitations, and directions for future research.

5.1 Effectiveness of the methodology

VEDANet has demonstrated remarkable effectiveness in enhancing visual emotion analysis across various datasets, including both controlled and in-the-wild settings. The architecture's ability to learn deep, nuanced representations of facial emotions results in significantly improved accuracy rates in emotion detection tasks. The OTO layer, in particular, addresses the challenge of capturing the multiplicity of human emotions expressed even in a single instance, enabling a more refined and accurate emotion analysis.

5.2 Architectural innovations

The architectural innovations of VEDANet, especially the OTO layer, play a pivotal role in its success. This layer employs a sophisticated mechanism to disentangle the intricate

patterns of emotional expressions, allowing for a more precise classification of complex emotional states. It acts by expanding the conventional one-to-one mapping paradigm of emotion recognition into a more flexible one-to-many framework, thus capturing a broader range of emotional nuances from visual cues.

Comparative analysis with existing benchmarks indicates that VEDANet outperforms prior models in various aspects, particularly in terms of accuracy and robustness. While traditional models often struggle with the subtlety and variability of emotional expressions, especially in less controlled environments, VEDANet's architecture, leveraging the OTO layer, demonstrates superior performance. It effectively reduces the gap between controlled laboratory settings and real-world applications, a longstanding challenge in the field of emotion recognition.

5.3 Potential applications

The implications of VEDANet's advancements are vast, impacting several fields where emotion recognition is crucial. In healthcare, for example, VEDANet could enhance patient care by providing real-time emotional analysis, aiding in the diagnosis and treatment of psychological conditions. In the automotive industry, emotion recognition can improve safety by monitoring drivers' states and preventing accidents caused by drowsiness or stress. Moreover, in customer service, understanding customer emotions through visual cues can lead to more personalized and effective interactions.

5.4 Healthcare and mental health monitoring

As the work uses cognitive neural assessment of image sequences to determine human behavior [49]. Exploring another avenue is an interesting approach to analyze bosom malignancy by assessment of mammograms using machine learning [50]. Further [51] study proposes a game for development of emotional and social skills in differently abled autistic people. VEDANet can be employed to develop tools for psychologists and psychiatrists to monitor patients' emotional states, providing valuable insights into their mental health and progress over time. In telehealth platforms, it can assist therapists by offering an additional layer of emotional context during virtual sessions, enabling better assessment and treatment plans for conditions like depression or anxiety.

5.5 Education and e-learning

In the educational sector, VEDANet can be integrated into e-learning platforms to gauge students' emotional responses to different learning materials. This emotional feedback can help educators tailor content delivery, ensuring it resonates positively with students, thereby improving engagement and learning outcomes. Furthermore, it can assist in identifying students who may be struggling with stress or frustration, allowing for timely intervention.

5.6 Customer service and market research

VEDANet has the potential to revolutionize customer service by enabling more empathetic interactions between service agents and customers. By analyzing customers' emotional

states during service interactions, companies can adjust their communication strategies in real time, leading to improved customer satisfaction. In market research, VEDANet can analyze consumer reactions to products or advertisements, providing valuable insights into consumer preferences and behaviors.

5.7 Entertainment and media

In the entertainment industry, VEDANet can be used to test audiences' emotional responses to movies, TV shows, or video games, helping creators fine-tune their content for maximum emotional impact. Additionally, it could enhance interactive media experiences by adjusting content dynamically based on the user's emotional state, creating more immersive and personalized entertainment experiences.

5.8 Smart home and IoT devices

Smart home devices and IoT systems can leverage VEDANet to become more responsive to users' emotional and psychological needs. For example, a smart home system could adjust lighting, music, and temperature based on the occupants' mood, creating a more comfortable and supportive living environment.

5.9 Automotive industry

In the automotive sector, VEDANet can enhance driver safety by monitoring drivers' emotional states and alertness levels, providing warnings or taking corrective actions (like limiting the vehicle's speed) when necessary. It can also improve the overall driving experience by adapting vehicle settings and entertainment options according to the driver's mood.

5.10 Limitations and future research directions

While VEDANet introduces significant advancements in visual emotion analysis, it is not without limitations. One of the primary concerns is dataset biases, which can impact the model's ability to generalize across different demographics and cultural contexts. Additionally, while VEDANet excels in analyzing visual data, human emotions are complex and can be influenced by a multitude of factors beyond facial expressions.

Future research could explore integrating multimodal data (such as audio, textual, and physiological signals) with VEDANet to create a more comprehensive understanding of emotions. This could involve developing hybrid models that can analyze and interpret signals from multiple sources simultaneously. Another promising direction is exploring unsupervised or semi-supervised learning approaches to overcome the challenges of dataset biases and limited labeled data, potentially enabling VEDANet to learn from a broader range of emotional expressions and contexts.

By addressing these limitations and exploring these future research directions, VEDANet can evolve into an even more powerful tool for emotion recognition, with widespread applications across various domains where understanding human emotions is vital.

VEDANet represents a significant leap forward in the field of visual emotion analysis, offering innovative solutions to longstanding challenges. By emphasizing architectural innovations like the OTO layer, VEDANet not only achieves high accuracies across

varied datasets but also opens new avenues for real-world applications of emotion recognition technology. While acknowledging its current limitations, the proposed future research directions highlight the potential for further advancements, paving the way for more accurate, adaptable, and comprehensive emotion recognition systems.

6 Conclusion and Possible Future Directions

In this work, we introduce VEDANet, a robust and efficient architecture designed for the nuanced task of facial expression detection and analysis in both still images and video sequences. Leveraging the power of Transfer Learning, our approach excels in automatically extracting distinctive facial features, crucial for accurate emotion recognition. The integration of hybrid pre-trained Deep Convolutional Neural Network (DCNN) models significantly enhances our system's performance, as evidenced by our comprehensive experimental analysis across several esteemed datasets, including FER2013, AffectNet, Google FEC, and Yale emotion datasets.

The VEDANet architecture sets a new benchmark in emotion recognition accuracy, achieving an impressive 87.3% on AffectNet, 92.7% on Google FEC, 95.0% on Yale, 90.5% on FER2013, and 93.4% in real-time video analysis. These results are attained through the strategic application of the Top Optimization Layer (OTO-Layer), which meticulously fine-tunes the hyperparameters of each pre-trained model to focus on the salient facial features present in each image. This meticulous attention to detail ensures that VEDANet not only outperforms existing state-of-the-art techniques but also addresses some of the limitations observed in previous models, such as misclassifications under normal conditions and challenges in interpreting complex facial expressions.

The practical implications of our study extend far beyond the laboratory, offering valuable contributions to various commercial and industrial applications. For instance, the deployment of VEDANet in patient monitoring systems within healthcare facilities can provide caregivers with real-time, non-intrusive insights into patient emotional well-being, enhancing patient care and support. Similarly, in the realm of security and surveillance, VEDANet's ability to accurately analyze facial expressions in real-time can enhance threat detection and response mechanisms, ensuring safer public and private spaces.

Furthermore, the foundation laid by our study in facial emotion detection and analysis opens up exciting avenues for incorporating additional modalities, such as body language and auditory signals, to create a more holistic approach to emotion recognition. Such multimodal emotion detection systems could revolutionize a myriad of industries, from enhancing user experience in consumer electronics to providing deeper audience insights in marketing and entertainment.

In conclusion, VEDANet represents a significant advancement in the field of emotion recognition, combining high accuracy with real-world applicability. Our research not only sets new standards in facial expression analysis but also paves the way for future innovations in multimodal emotion detection, promising to unlock new dimensions in human-computer interaction.

Data availability The datasets generated during and/or analysed during the current study are available in the respective repositories as followed: 1. Affectnet [http://mohammadmahoor.com/affectnet/]. 2. FER2013 [https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data] 3. GoogleFEC [https://research.google/tools/datasets/google-facial-expression/]. 4. Extended Yale dataset [https://computervisiononline.com/dataset/1105138686]

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Sariyanidi E, Gunes H, Cavallaro A (2014) Automatic analysis of facial affect: A survey of registration, representation, and recognition. IEEE Trans Pattern Anal Mach Intell 37(6):1113–1133
2. Anagnostopoulos CN, Iliou T, Giannoukos I (2015) Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011. Artif Intell Rev 43(2):155–177
3. Marechal C, Mikolajewski D, Tyburek K, Prokopowicz P, Bougueroua L, Ancourt C, Wegrzyn- Wolska K (2019) Survey on AI-Based Multimodal Methods for Emotion Detection. High-Perform Model Simul Big Data Appl 11400:307–324
4. Alkawaz MH, Mohamad D, Basori AH, Saba T (2015) Blend shape interpolation and FACS for realistic avatar. 3D Research 6(1):1–10
5. Shan G, Shan C, Gong S, McOwan PW (2009) Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image Vis Comput 27(6):803–816
6. Zhang S, Li L, Zhao Z (2012) Facial expression recognition based on Gabor wavelets and sparse representation. In: 2012 IEEE 11th International conference on signal processing. IEEE, vol 2, pp 816–819
7. Rouast PV, Adam MT, Chiong R (2019) Deep learning for human affect recognition: Insights and new developments. IEEE Trans Affect Comput 12(2):524–543
8. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, pp 7132–7141. https://doi.org/10.1109/CVPR.2018.00745
9. Liu ZT, Li SH, Cao WH, Li DY, Hao M, Zhang R (2019) Combining 2D gabor and local binary pattern for facial expression recognition using extreme learning machine. J Adv Comput Intell Intell Inf 23(3):444–455
10. Deeb H, Sarangi A, Mishra D, Sarangi SK (2022) Human facial emotion recognition using improved black hole based extreme learning machine. Multimed Tools Appl 1–24
11. B+. Available online: https://computervisiononline.com/dataset/1105138686. Accessed 29 Nov 2017
12. Vemulapalli R, Agarwala A (2019) A compact embedding for facial expression similarity. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 5683–5692
13. Mollahosseini A, Hasani B, Mahoor MH (2017) Affectnet: A database for facial expression, valence, and arousal computing in the wild. IEEE Trans Affect Comput 10(1):18–31
14. Goodfellow IJ, Erhan D, Carrier PL, Courville A, Mirza M, Hamner B, Bengio Y (2013) Challenges in representation learning: A report on three machine learning contests. In: International sconference on neural information processing. Springer, Berlin, Heidelberg, pp 117–124
15. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2818–2826
16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
17. Huang G, Liu Z, vd Maaten L, Weinberger KQ (2017) In Densely connected convolutional networks. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 2261–2269
18. Chollet F (2017) Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1251–1258
19. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, Hasan M, Van Essen BC, Awwal AA, Asari VK (2019) A state-of-the-art survey on deep learning theory and architectures. Electronics 8(3):292
20. Sahu M, Dash R (2021) A survey on deep learning: convolution neural network (CNN). In: Intelligent and cloud computing. Springer, Singapore, pp 317–325
21. Mollahosseini A, Chan D, Mahoor MH (2016) Going deeper in facial expression recognition using deep neural networks. In: 2016 IEEE Winter conference on applications of computer vision (WACV). IEEE, pp 1–10
22. Zhao X, Shi X, Zhang S (2015) Facial expression recognition via deep learning. IETE Tech Rev 32(5):347–355

23. Li J, Huang S, Zhang X, Fu X, Chang CC, Tang Z, Luo Z (2018) Facial expression recognition by transfer learning for small datasets. In: International conference on security with intelligent computing and big-data services. Springer, Cham, pp 756–770
24. Bendjillali RI, Beladgham M, Merit K, Taleb-Ahmed A (2019) Improved facial expression recognition based on DWT feature for deep CNN. Electronics 8(3):324
25. Ngoc QT, Lee S, Song BC (2020) Facial landmark-based emotion recognition via directed graph neural network. Electronics 9(5):764
26. Pranav E, Kamal S, Chandran CS, Supriya MH (2020) Facial emotion recognition using deep convolutional neural network. In: 2020 6th International conference on advanced computing and communication Systems (ICACCS). IEEE, pp 317–320
27. Khan A, Sohail A, Zahoora U, Qureshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 53(8):5455–5516
28. Hochreiter S, Bengio Y, Frasconi P, Schmidhuber J (2001) Gradient flow in recurrent nets: the difficulty of learning long-term dependencies
29. He K, Zhang X, Ren S, & Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
30. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
31. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
32. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
33. Zhou F, Kong S, Fowlkes CC, Chen T, Lei B (2020) Fine-grained facial expression analysis using dimensional emotion model. Neurocomputing 392:38–49
34. Kuruvayil S, Palaniswamy S (2021) Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning. J King Saud Univ–Comput Inf Sci
35. Wang C, Lu K, Xue J, Yan Y (2019) Dense attention network for facial expression recognition in the wild. In: Proceedings of the ACM multimedia asia, pp 1–6
36. Ahadit AB, Jatoth RK (2022) A novel multi-feature fusion deep neural network using HOG and VGG-Face for facial expression classification. Mach Vis Appl 33(4):1–23
37. Schoneveld L, Othmani A, Abdelkawy H (2021) Leveraging recent advances in deep learning for audio-visual emotion recognition. Pattern Recognit Lett 146:1–7
38. Rescigno M, Spezialetti M, Rossi S (2020) Personalized models for facial emotion recognition through transfer learning. Multimed Tools Appl 79(47):35811–35828
39. Li Y, Zeng J, Shan S, Chen X (2018) Occlusion aware facial expression recognition using CNN with attention mechanism. IEEE Trans Image Process 28(5):2439–2450
40. Minaee S, Minaei M, Abdolrashidi A (2021) Deep-emotion: Facial expression recognition using attentional convolutional network. Sensors 21(9):3046
41. Pecoraro R, Basile V, Bono V (2022) Local multi-head channel self-attention for facial expression recognition. Information 13(9):419
42. Lin DT, Pan DC (2009) Integrating a mixed-feature model and multiclass support vector machine for facial expression recognition. Integr Comput Aid Eng 16(1):61–74
43. Saroop, A., Ghugare, P., Mathamsetty, S., & Vasani, V. (2021). Facial Emotion Recognition: A multi-task approach using deep learning. arXiv preprint arXiv:2110.15028.
44. Acharya D, Huang Z, Pani Paudel D, Van Gool L (2018) Covariance pooling for facial expression recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 367–374
45. Greco A, Strisciuglio N, Vento M, Vigilante V (2023) Benchmarking deep networks for facial emotion recognition in the wild. Multimed Tools Appl 82(8):11189–11220
46. Amal VS, Suresh S, Deepa G (2022) Real-time emotion recognition from facial expressions using convolutional neural network with Fer2013 dataset. In: Ubiquitous intelligent systems: Proceedings of ICUIS 2021. Springer Singapore, pp 541–551
47. Lv JJ, Shao XH, Huang JS, Zhou XD, Zhou X (2017) Data augmentation for face recognition. Neurocomputing 230:184–196
48. Glorot X, Bengio Y (2010) Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, pp 249–256

49. Parashar J, Kushwah VS, Rai M (2023) Determination human behavior prediction supported by cognitive computing-based neural network. In: Soft Computing: Theories and Applications: Proceedings of SoCTA 2022. Springer Nature Singapore, Singapore, pp 431–441
50. Kushwah VS, Saxena A, Pahariya JS, Goyal SK (2021) Support vector machine technique to prognosis breast cancer. In: Soft Computing: Theories and Applications: Proceedings of SoCTA 2020, vol 2. Springer Singapore, pp 339–351
51. Keating CT, Ichijo E, Cook JL (2023) Autistic adults exhibit highly precise representations of others' emotions but a reduced influence of emotion representations on emotion recognition accuracy. Sci Rep 13:11875. https://doi.org/10.1038/s41598-023-39070-0

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

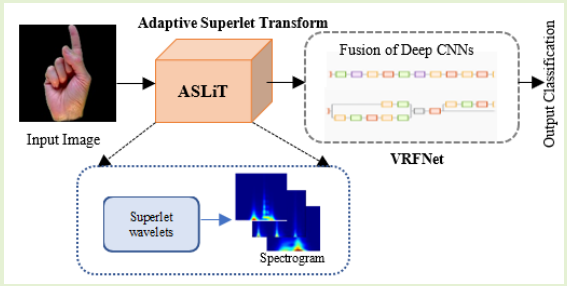
This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

IEEE Sensors Council

VRFNet-ASLiT: Fused Deep CNN and Adaptive Super Resolution Transform Based Hand Gesture Recognition

Roli Kushwaha, Manjeet Kumar*, Dinesh Kumar

***Abstract*— Hand gesture recognition plays a vital role in human-computer interaction, offering a natural and intuitive means of communication. However, vision-based hand gesture recognition is sensitive to variations in lighting conditions and may face challenges in accurately distinguishing between similar gestures or subtle variations in hand. To address this concern, a hybrid deep Convolutional Neural Network (CNN) model which is the fusion of VGG19 and ResNet50 ($VRFNet$) is proposed to encapsulate the intricacies of subtle variations or infrequent gestures. The proposed method incorporates the Adaptive Super Resolution Transform ($ASLiT$) technique to generate the high-resolution time-frequency representation of input images and $VRFNet$ effectively classify the different hand gestures. The proposed method is tested on LeapGestRecog dataset, Modified National Institute of Standards and Technology (MNIST) digits dataset and American Sign Language (ASL) digits and provides a recognition accuracy of 99.97%, 99.6% and 98.8%, respectively. In addition to the proposed $VRFNet$ model, the comparison of the performance with other two other models: AlexNet, VGG19 is also presented. The results reveal that $VRFNet$ model exhibits exceptional performance in hand gesture recognition tasks with improved precision of 99.83%, 99.64%, 99.05%, recall of 99.86%, 99.62%, 99.85% and F1 score of 98.85%, 99.63%, 98.96%, respectively, across the LeapGestRecog, MNIST, and ASL digits dataset. These results underscore the model's remarkable accuracy in recognizing hand gestures within diverse datasets. The proposed $VRFNet$ model is also competitively, comparable to other state-of-the-art hand gesture recognition techniques. Furthermore, an ablation study is conducted to analyze the impact of individual models for hand gesture recognition.**



***Index Terms*— Adaptive Superlet Transform, AlexNet, VGG19, ResNet50, Hand Gesture Recognition, Human Computer Interaction, Multiclass classification.**

I. Introduction

GESTURES play a crucial role in facilitating communication between humans and machines [1]. It allows users to convey commands, instructions, or intentions. The most used gestures of body for communication with machines is through intuitive hand gestures which enables a more natural and expressive form of communication with technology. Effective human-computer communication (HCI) can be achieved through various techniques such as machine learning [2], deep learning [3], reinforcement learning [4]. In comparison to other approaches, traditional machine learning relies on human feature engineering, limiting its ability to detect complex patterns in hand gesture images [2]. Reinforcement learning often relies on trial-and-error exploration, which is inefficient for high-dimensional tasks like image analysis [4]. Among these approaches, deep learning offers distinct advantage of automatically learn and extract complex patterns from data, making it an appropriate choice for the research. Convolutional Neural Networks (CNNs) are mostly used deep learning models particularly well-suited for image-based tasks, making them highly applicable for analyzing hand gesture images [5]. The recognition of hand gestures can be done using image-based analysis method or surface Electromyography signals (sEMG) based analysis method. The image-based analysis of hand gestures is preferred over employing sEMG due to the unique advantages of visual information, resistance to variability, natural representation, and the accessibility of large-scale datasets and provide a more thorough and robust method for hand gesture identification [6]. These advantages make image-based analysis a preferred choice over sEMG for many applications in human-computer interaction and gesture recognition.

Image-based analysis methods like deep learning have made it possible to extract significant features and patterns from images of hand gestures. In case of dynamic hand gestures, temporal variation plays a cruical role, thereby integrating image-based analysis with transform-based techniques to

[1] The authors are with the Department of Electronics and Communication Engineering, Delhi Technological University, Delhi, 110042, India.

(e-mail: rolikushwaha.bbd@gmail.com; manjeetchhillar@gmail.com; dineshkumar@dtu.ac.in).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

examine the frequency characteristics of hand that enables the detection of temporal variations and dynamic nature of gestures [7]. Additionally, the time-frequency representation obtained through transform-based techniques [7]-[9] provides valuable insights into the temporal dynamics and rhythm of the gestures.

Existing transform-based techniques, such as Short-Time Fourier Transform (STFT) [7], which provides fixed time and frequency resolution, Adaptive Wavelet Transform (AWT) [9], adapting wavelet basis functions but potentially sacrificing resolution due to its susceptibility to noise and computational complexity, and Continuous Wavelet Transform (CWT) [8], offering superior time-frequency resolution through continuous wavelet scale variation but sacrificing frequency resolution due to variable scaling, often do not meet the high-resolution in both time and frequency simultaneously. To address this issue, a Super Resolution Transform (Superlet or SLT) was introduced by Moca et al. [10] [11] which provides high-resolution in both time-domain and frequency-domain. Thus, in order to obtain outstanding recognition accuracy of hand gestures, the proposed methodology incorporates hybrid deep CNN called $VRFNet$ including Adaptive Super Resolution Transform (ASLT) for hand gesture images called $ASLiT$ which offer valuable insights into the temporal dynamics and rhythm of the gestures.

The main contributions can be summarized as follows:

1) A hybrid hand gesture classification network $VRFNet$ is proposed. $VRFNet$ is a fusion of two deep CNN models namely ResNet50 and VGG19. The proposed hybrid model has the ability to extract high-level discriminative features from images which produces more precise predictions.
2) The proposed model leverages $ASLiT$ that transform input images into a time-frequency representation, enabling the model to achieve outperforming classification results of hand gestures.
3) Extensive experiments are conducted on multimodal dataset: LeapGestRecog (infrared images), MNIST digits (gray-scale images) and American Sign Language (ASL) digits (RGB images) to validate the efficiency and efficacy of the proposed model.
4) Further, an ablation study is also carried that justifies the performance effectiveness of the proposed methodology.

The organization of the rest of the paper is as follows: Section II discusses about the different state-of-the-art models for $ASLiT$ and hand-gesture recognition. Section III introduces proposed methodology including brief of $ASLiT$ and proposed $VRFNet$ deep CNN model-based recognition. The results of the simulations and discussion are presented in Section IV. Lastly, Section V presents the conclusions and future scope of this paper.

II. Related Works

This section provides the details of the previous works done for hand gesture recognition and ASLT in signal processing.

A. Hand Gesture Recognition

In this section, the related works that have focused on hand gesture recognition, specifically considering the utilization of deep CNN models are explained. Park et al. [12] employed a two-stage serial CNN model and a double parallel CNN model for the classification of hand gestures and two-dimensional fast Fourier transform for converting time-domain hand gestures to frequency-domain, resulting in a faster and more accurate model. Damaneh et al. [13], presented a CNN, Gabor filter, and Oriented Fast and Rotated Brief (ORB) feature descriptor based static hand gesture recognition network by merging extracted features into one feature vector resulting in enhanced resistance to uncertainties like rotation and ambiguity in hand gestures. In [14], researchers introduced HyFinet, a lightweight portable network with 2.3 million parameters, for hand gesture classification. It utilized multi-scale refined edge extraction and hybrid feature attention blocks, enhancing recognition accuracy with lower computational cost. Static hand gesture recognition, utilized EMG signals from three sensors by extracting time domain features and employing linear discriminant analysis is presented in [15]. For sEMG-based gesture recognition, a recursive neural network (RNN) architecture with six different types of images as an input and an attention-based CNN were discussed in [16]. Numerous studies also suggest gesture classification methods by converting sEMG signal to 2-D and then applying deep learning models for classification [17]-[19]. The majority of recent research on hand gesture recognition focuses on identifying static hand gestures. However, dynamic gestures are more natural and user-friendly for being employed as an alternative input modality in HCI applications. In [20], Tripathi et al. presented a dynamic hand gesture recognition model based on 2D CNN using EMG signals which are transformed into time–frequency representations using STFT and CWT. A multi-stream residual network was proposed by Yang et al. [21] for the classification of six dynamic gestures involving the complete hand movement based on sEMG data. Various researches are based on multimodal fusion for hand gesture recognition [22] [23]. In [22], gesture recognition based on Multimodal Fusion Hierarchical Self-Attention Network was employed that combines features from different modalities with skeleton features using self-attention blocks at multiple hierarchical levels. Rastgoo et al. [23], presented transformer based zero-shot dynamic hand gesture recognition by fusing deep features with skeleton based features extracted by 3D CNN.

B. Adaptive Superlet (ASLT) in Signal Processing

ASLT is widely used in different areas of signal processing; biosignals like Electrocardiogram (ECG) and Photoplethysmogram (PPG) in medical field [24]-[30], in industries like faults in bearings of motor in the form of vibration signals [31]-[34]. Tripathi et al. [24] utilized the ASLT to convert the 1-D ECG signal into a 2-D time-frequency spectrogram, which served as input to deep learning pre trained classifiers. In [25], heart rate (HR) monitoring was presented by Pankaj et al. using PPG signal by utilizing a multiorder super resolution transform (SLT) to separate motion artifacts from the acquired PPG signal resulting in detection of high-intensity lobes in the spectro-temporal representation of the SLT spectrogram. Bhatt et al. [26], introduced a method to identify Parkinson's disease by combining the Superlet Transform with Deep Neural Networks in order to get high-resolution time-frequency spectrogram. Tripathi et al. [27], presented a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

technique for identifying seizure and non-seizure events using the superlet transform and a deep convolutional neural network. VGG-19. In [31], authors provided an ASLT methodology for fault by surpassing the s-transform in terms of both time and frequency resolution for fault identification in power systems. In [32], Mitra et al. introduced a fusion method incorporating the time-frequency characteristics of vibration signals using wavelets' super resolution property and a 2-D-CNN. Yi et al. [33], exhibited the novel method called High-order Synchrosqueezing Superlets Transform by demonstrating time frequency representation of vibration signals in mechanical transmission components under variable speed conditions.

Based on the notable performance of ASLT in signal processing domains, it provides a strong foundation for exploring its potential in image processing. Therefore, the proposed research incorporates the *ASLiT* in hand gesture recognition approach and aims to achieve higher recognition accuracy and performance.

III. Proposed Methodology

This section discusses the proposed methodology which utilizes *ASLiT* for images and an improved version of deep CNN. The proposed methodology aims to enhance the quality of images by using *ASLiT* and generating high-resolution spectrogram by effectively capturing fine-grained features. The classification of different gestures is performed by an improved version of CNN (*VRFNet*), thereby outperforming with existing hand gesture recognition system in terms of recognition accuracy.

A. Adaptive Superlet Transform for Images (ASLiT)

ASLiT is a modified version of Morlet transform, which provides high-resolution in localization of both time and frequency simultaneously unlike STFT [7], AWT [9] and CWT [8]. STFT segments signals for Fourier Transform, but fixed window sizes compromise capturing both high and low frequencies simultaneously, affecting frequency-temporal resolution. AWT adapts wavelet basis functions based on signal characteristics, offering flexibility but potentially sacrificing resolution due to its susceptibility to noise and computational complexity, making it less ideal for capturing rapidly changing features whereas CWT offers variable time-frequency resolution across frequencies, providing good temporal resolution at higher frequencies by compressing or dilating a mother wavelet based on frequency variations but sacrificing frequency resolution. To address this concern, SLT is introduced by Moca et al. [10] [11] to attain precise localization of both time and frequency concurrently. SLT utilizes a set of wavelets with increasingly constrained bandwidth. The use of multiple wavelets enables the localization of oscillations in time-frequency plane is improved than using single wavelet. SLT employs Morlet wavelet that is a complex-valued wavelet, derived from a Gaussian function modulated by a complex sine wave, thus making it well-suited for analysing non-stationary signals with time-varying frequencies.

The mathematical expression of mother wavelet in Morlet wavelet is shown in equation (1).

$$\psi_{fc,r}(t) = \frac{1}{T_r\sqrt{2\pi}} e^{-\frac{t^2}{2T_r{}^2}} e^{j2\pi f_c t} \quad (1)$$

where, fc is central frequency, T_r is time spread parameter (in seconds) and T_r which depends on number of base cycles r, standard deviation (SD) and fc as given by eq. 2. It determines the width of the time-domain representation of the wavelet. A larger T_r leads to a broader time-domain representation, which can capture lower-frequency components.

$$T_r = \frac{r}{SD \times f_c} \quad (2)$$

In addition to the 1D Morlet transform, there is also a 2D Morlet transform that extends the concept to analyze 2D signals such as images [12]. By convolving an image with the 2D Morlet wavelet at different scales and orientations, results in a time-frequency representation of the image. The expression of the 2D Morlet transform is shown:

$$\psi_{f_c,r}(x,y) = \frac{1}{T_r\sqrt{2\pi}} e^{\left(-\frac{x^2+y^2}{2T_r^2}\right)} e^{j2\pi f_c(x+y)} \quad (3)$$

where, $\psi_{f_c,r}(x,y)$ represents the 2D Morlet wavelet at spatial coordinates (x,y), fc is the central frequency of the wavelet in the x and y directions corresponding to the pixel positions, respectively, (x^2+y^2) shows how the pixel values of an image are affected by the Gaussian envelope and $(x+y)$ shows the spatial variation in the sinusoidal modulation across the image.

The Morlet wavelet's fixed shape and size limit its ability to capture diverse features in complex images, leading to less accurate localization and representation. It also struggles to effectively capture frequency content across image regions [14] [34], hindering precise time-frequency localization and fine-scale temporal variation capture. To overcome this limitation, SLT approach can be employed [13]. This approach involves using multiple Morlet wavelets with varying scales to analyze the image at different frequency resolutions. By adaptively selecting the appropriate scale for each image region, the spectrogram can better capture the frequency content and temporal variations across different scales, results in generating high-resolution spectrograms [11]. The mathematical expression of SLT based time-frequency analysis of images is:

$$SLiT_{f_c,M} = \{\psi_{f_c,r_x r_y} | r_x = r_1, r_2, \ldots r_M \; and \; r_y = r_1, r_2, \ldots r_M \quad (4)$$

where, $SLiT_{fc,M}$ represents the SLT that is obtained by set of 2D wavelets of images using order (number of wavelets) of M with varying the scale parameters (number of cycles for each wavelet) r_x and r_y within a range of $r_1, r_2, \ldots r_M$, $\psi_{fc,r_x r_y}$ represents the 2D wavelet at scale (r_x, r_y) and frequency f_c, r_x denotes the scale parameter in the x direction, while r_y denotes the scale parameter in the y direction, $r_1, r_2, \ldots r_M$ represent the range of scales (number of cycles) to be considered in both the x and y spatial coordinates.

$SLiT_{f_c,M}$ (SLiT) is also a measure that represents the geometric mean of the individual wavelet responses in the set defined by eq. (5). In other words, the SLiT combines the responses of multiple wavelets $R[SLiT_{fc,M}]$ at different scales using a geometric mean [11].

$$R\left[SLiT_{f_c,M}\right] = \sqrt[M]{\prod_{i=1}^{M} R[\psi_{f_c, r_{x_i}, r_{y_i}}]} \quad (5)$$

$R[\psi_{f_c, r_{x_i}, r_{y_i}}]$ is the response of i_{th} wavelet at frequency f_c and

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

scale r_{x_i}, r_{y_i} is evaluated at the position (x_i, y_i) in the image I.

$$R\left[\psi_{fc,r_{x_i},r_{y_i}}\right] = \sqrt{2} \cdot I(x,y) * \psi_{fc,r_{x_i},r_{y_i}} \quad (6)$$

where $I(x,y)$ is an image and $*$ is the complex convolution operator. By calculating $R[SLiT_{fc,M}]$, a single value that represents the combined response of the SLiT across multiple scales and positions, summarizing the frequency content and spatial variations within the image at the specified frequency *fc* is obtained.

There are two ways to select number of cycles for each wavelet (r_{x_i}, r_{y_i}) explained in [13] viz. multiplicatively or additively where, $(r_{x_i}, r_{y_i}) = i \,.\, r_{x_1} r_{y_1}$ and $(r_{x_i}, r_{y_i}) = \{(r_{x_1} + i), (r_{y_1} + i)\} - 1$ for $i = 2, \ldots, M$) respectively. If the order of SLiT is fixed, then it can cause increase of wavelet bandwidth while increasing frequency. To recompense this issue, Adaptive SLiT ($ASLiT$) is designed which can adjust its order (M) to the central frequency (f_c). It works with low order for low frequencies and as frequency increases, it increases its order to provide enhanced time-frequency representations [11]. The process of increasing order is dependent on frequency as given by eq. (7).

$$ASLiT_{f_c} = SLiT_{f_c,M} | M = N(f_c) \quad (7)$$

where, $N(f_c)$ represents a function that increases in a consistent manner as the central frequency (f_c) increases. One straightforward approach is to select the order in a linear fashion, where the order increases proportionally with the central frequency given by eq. 8

$$N(f_c) = M_{min} + \left[(M_{max} - M_{min}).\frac{f_c - f_{c_{min}}}{f_{c_{max}} - f_{c_{min}}}\right] \quad (8)$$

where, M_{min}, M_{max} are the orders corresponding to smallest central frequency $f_{c_{min}}$ and largest central frequency $f_{c_{max}}$. $ASLiT$ generates high-resolution spectrogram for wide range of frequencies, thereby the proposed methodology utilizes the power of $ASLiT$ for hand gesture recognition.

B. Hand gesture Classification Network

The proposed methodology for hand gesture classification utilizes the power of $ASLiT$ that is employed to enhance the time-frequency representation of input images. This enhancement step plays a crucial role in mitigating the negative effects of low-resolution images on the accuracy of hand gesture recognition. Additionally, we incorporate an improved version of CNN into the proposed methodology which is hybrid model consisting of fusion of VGG19 and ResNet50. By concatenating the features learned by different models, the hybrid model can benefit from the strengths of each model and potentially achieve better overall performance. This enhanced CNN architecture is specifically designed to handle the high-resolution spectrogram generated by the $ASLiT$. The block diagram of the proposed methodology is shown in Fig. 1. The model inputs the infrared, gray-scale and RGB images from datasets named LeapGestRecog, MNIST digits and ASL digits dataset, respectively. The images are transformed into time-frequency representation using $ASLiT$ by taking the geometric mean of the responses of multiple wavelets as shown in eq. (5).

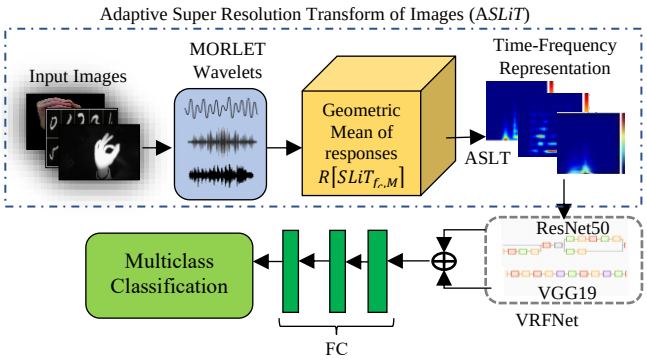


Fig. 1. Block diagram representation of the proposed $VRFNet$ based hand gesture recognition model.

The process of obtaining geometric mean results in generating high-resolution spectrograms.

The spectrograms of each dataset are saved which are further used to train the deep learning model $VRFNet$. The $VRFNet$ is the fusion of VGG19 and ResNet50 after concatenation of fully connected layers of both models. The output of the final max pooling layer of the VGG19 architecture is selected and the selected features are then fed into a new 1×1024 Fully Connected (FC) layer with ReLU activation function. Similarly, the output of the final average pooling layer of the ResNet50 architecture is taken and the features are fed into a separate 1×1024 FC layer with ReLU activation. To fuse the features generated from the VGG19 and ResNet50 architectures, a new 1×2048 FC layer is added. Additionally, a 1×512 FC layer is employed to extract more prominent features. Finally, the final classification task is performed using a 1×10 FC layer with SoftMax activation function as shown in Fig. 2.

Different hyperparameters selected for the proposed model are optimizer 'rmsprop', learning rate '0.00001', epochs '40', batch size '32', loss function 'categorical cross entropy', validation frequency '20'. The categorical cross entropy loss measures the losses for multiclass classification:

$$CEL = -\frac{1}{N_i}\sum_{n=1}^{N_i}\sum_{c=1}^{C_l}(T'_{nc}\, ln(I_{nc}) + (1 - T'_{nc})\, ln(1 - I_{nc})) \quad (9)$$

where N, T', I, and C_l represents number of observations, target input image set, number of observations and number of classes respectively. The categorical cross entropy loss function reduces training and testing loss. The algorithm for the proposed model for classification of different gestures using $ASLiT$ and $VRFNet$ is presented as Algorithm 1.

C. Datasets Used

Three different datasets are used for evaluating the performance of the proposed model. These datasets contain variety of images (multimodal); infrared, grayscale and RGB images.

The 'LeapGestRecog' dataset [35] [36] is a collection of 10 distinct hand gestures generated with 20,000 infrared images of 640×240 pixels, taken with the Leap Motion sensor. The different gestures include; Palm, Fist, Fist Moved, Thumb, Index, OK, Palm_moved, Down, C letter, and L letter constitute the hand motions in the collection. The MNIST digit dataset [37] [38] consists of 70,000 grayscale images of 28×28 pixels for handwritten digits, with 60,000 images for training and 10,000 images for testing for class 0 to 9.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

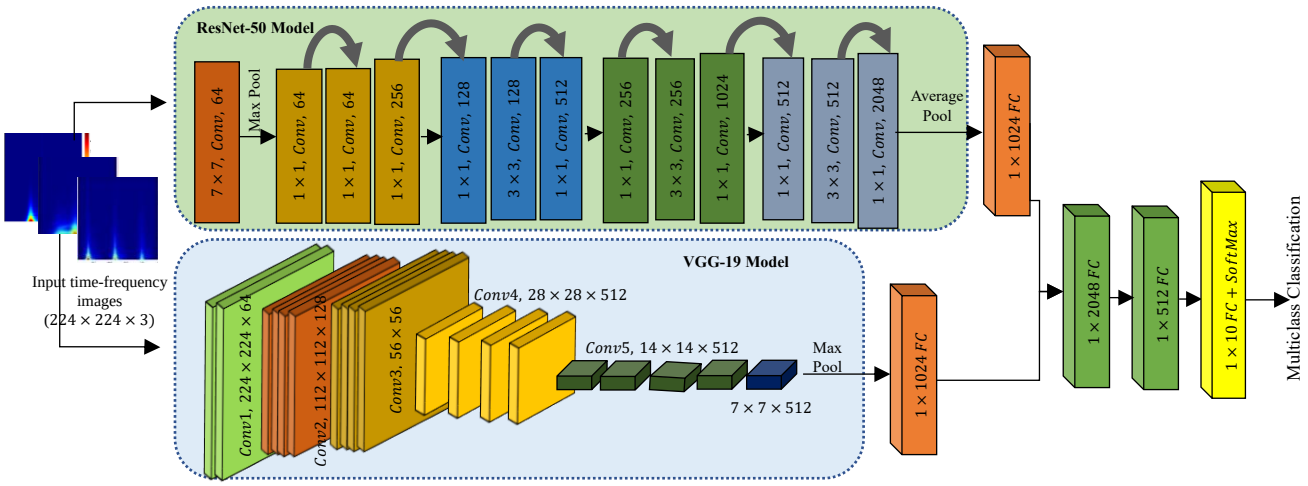


Fig. 2. Architecture of the proposed $VRFNet$ for hand gesture classification.

The ASL digit dataset [39] [40] is a collection of 5000 images of size 400 × 400 pixels that represent the numbers 0 to 9 in ASL. Some sample images of each dataset are shown in Fig. 3.

In order to visualize the datasets, t-SNE plots are presented in Fig. 4 (a), (b), and (c) for the three datasets. The t-SNE plots effectively capture the inherent structure and clustering of the datasets, enabling the observation of patterns and relationships among the data instances.

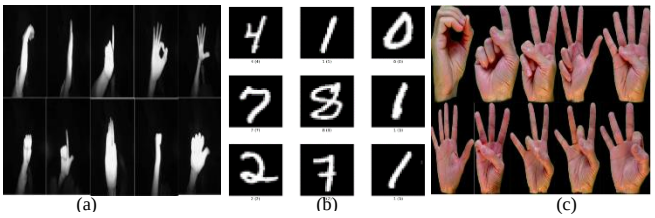


Fig. 3. Some sample images of (a) LeapGestRecog (b) MNIST digits and (c) ASL digits dataset.

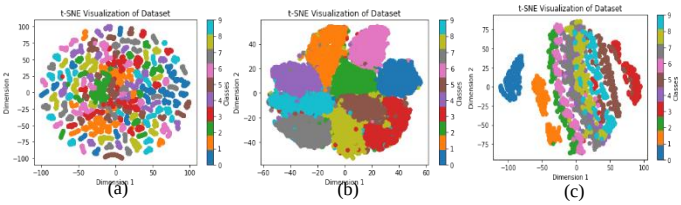


Fig. 4. t-SNE Visualization of (b) 'LeapGestRecog' (b) MNIST digits and (c) ASL digits dataset.

IV. Results and Discussions

In the proposed methodology, the utilization of the $ASLiT$ for generating high-resolution spectrogram followed by hybrid model $VRFNet$ for the classification of hand gesture. The high-resolution spectrograms of the datasets; LeapGestRecog, MNIST digits, and ASL digits datasets are shown in Fig. 5 (a), (b), (c), respectively.

The performance of the proposed methodology was evaluated by measuring the Accuracy (A_c), Precision (P_r) Recall (R_l) and F1 score (F_s) as shown in eqs. (10)-(13), respectively.

Algorithm 1 Proposed $ASLiT$ based hand gesture classification using $VRFNet$

Input: Images (Infrared, Gray, RGB), $I(x, y)$
Output: Classification of 10 classes (0-9 digits, 10 gestures), Accuracy (A_c), Precision (P_r), Recall (R_l), F1 score (F_s)

START

1. Convert input images into spectrograms using $ASLiT$:
2. Select mother wavelet:

$$\psi_{f_c,r}(x,y) = \frac{1}{T_r\sqrt{2\pi}} e^{\left(-\frac{x^2+y^2}{2T_r^2}\right)} e^{j2\pi f_c(x+y)}$$

3. Generation of spectrogram: $SLiT_{f_c,M} = \{\psi_{f_c,r_x r_y} | r_x = r_1, r_2, \ldots r_M \text{ and } r_y = r_1, r_2, \ldots r_M\}$
4. Multiplicative Spectrogram: $R\left[SLiT_{f_c,M}\right] = \sqrt[M]{\prod_{i=1}^{M} R[\psi_{f_c,r_{x_i},r_{y_i}}]}$

where, $R\left[\psi_{f_c,r_{x_i},r_{y_i}}\right] = \sqrt{2} \cdot I(x,y) * \psi_{f_c,r_{x_i},r_{y_i}}$

5. Apply $(ASLiT)$ with order (M) and central frequency (f_c) : $ASLiT_{f_c} = SLiT_{f_c,M} | M = N(f_c)$
6. Generated spectrogram by geometric mean of modified spectrogram of individual wavelets are used for training the $VRFNet$.
7. Prepare X_train, Y_train, X_test, Y_test
8. INITIALISE $VRFNet$
9. Train the $VRFNet$ by using X_train, Y_train
10. Compute the categorical cross-entropy loss (CEL) based on the predicted and true labels

$$CEL = -\frac{1}{N_i}\sum_{n=1}^{N_i}\sum_{c=1}^{C_l}(T'_{nc}\, ln(I_{nc}) + (1 - T'_{nc})\, ln(1 - I_{nc}))$$

11. Classification by X_test, Y_test.
12. Compute A_c, P_r, R_l, F_s

END

$$Accuracy\ (A_c) = \frac{(TP + TN)}{(TP + TN + FP + FN)} \times 100 \quad (10)$$

$$Precison\ (P_r) = \frac{TP}{TP+FP} \times 100 \quad (11)$$

$$Recall\ (R_l) = \frac{TP}{TP+FN} \times 100 \quad (12)$$

$$F1\ Score\ (F_s) = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \times 100 \quad (13)$$

where, TP denotes true positives, TN denotes true negatives, FP is false positive, FN is false negative outcomes.

The results revealed promising outcomes, with recognition accuracies of 99.97% with the LeapGestRecog, 99.6% using MNIST digits, and 98.8% with the ASL digits dataset as shown by the confusion matrix in Fig. 6 (a), (b), (c), respectively.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

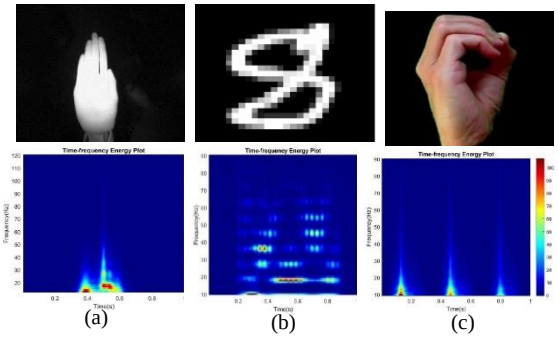


Fig. 5. Time-Frequency representation of some samples using $ASLiT$ of each dataset; (a) LeapGestRecog dataset, (b) MNIST digits dataset, and (c) ASL digits dataset for hand gesture recognition.

Furthermore, the three datasets are also tested with other deep learning models including AlexNet and VGG19. The accuracy of AlexNet was 95.29%, 93.9%, and 95.84% on the LeapGestRecog, MNIST digits, and ASL digits datasets, respectively. Meanwhile, VGG19 achieved accuracies of 96.83%, 96.41%, and 97.12% on the same datasets. The performance comparison of the proposed $VRFNet$ with AlexNet and VGG19 shows the outperforming results as reported in Table I.

The accuracies of different models at each epoch during training and testing process are shown by Box and Whisker plot to gain insights into their performance as depicted in Fig 7. It provides information about the distribution of accuracy per epoch (scaling factor of 1:4), central tendency, and variability of the accuracy scores. Also, the recognition accuracy and loss of the AlexNet, VGG19 and $VRFNet$ models using LeapGestRecog, MNIST digits, and ASL digits dataset are shown by Fig. 8 (a), (b) and (c), respectively.

TABLE I
COMPARISON OF THE PROPOSED MODEL $VRFNet$ WITH ALEXNET AND VGG19 USING DIFFERENT DATASETS FOR HAND GESTURE RECOGNITION

Dataset	Model	Evaluation Parameters (%)			
		A_c	P_r	R_l	F_S
LeapGestRecog	AlexNet	95.29	97.30	97.80	97.54
	VGG19	96.83	98.50	98.35	98.42
	VRFNet	**99.97**	**99.83**	**99.86**	**99.85**
MNIST digits	AlexNet	93.90	94.50	93.60	94.35
	VGG19	96.41	98.23	98.14	98.18
	VRFNet	**99.60**	**99.64**	**99.62**	**99.63**
ASL digits	AlexNet	95.84	96.90	96.20	96.54
	VGG19	97.12	98.04	99.01	98.52
	VRFNet	**98.80**	**99.05**	**98.85**	**98.96**

A. Comparative Study

The comprehensive comparison of the proposed methodology with state-of-the-art techniques that have been previously applied to the same datasets [35]-[46] as shown in Table II. Correspondingly, the comparison with existing methods using diverse dataset for hand gesture recognition [47]-[52] is given in Table III. The aim of this comparison is to evaluate the performance and effectiveness of the proposed approach vis-à-vis the state-of-the-art techniques applied for hand gesture recognition. From the comparison Table II, it is evident that the proposed methodology correctly distinguishes different hand gestures and provides a recognition accuracy of 99.97%, 99.6%

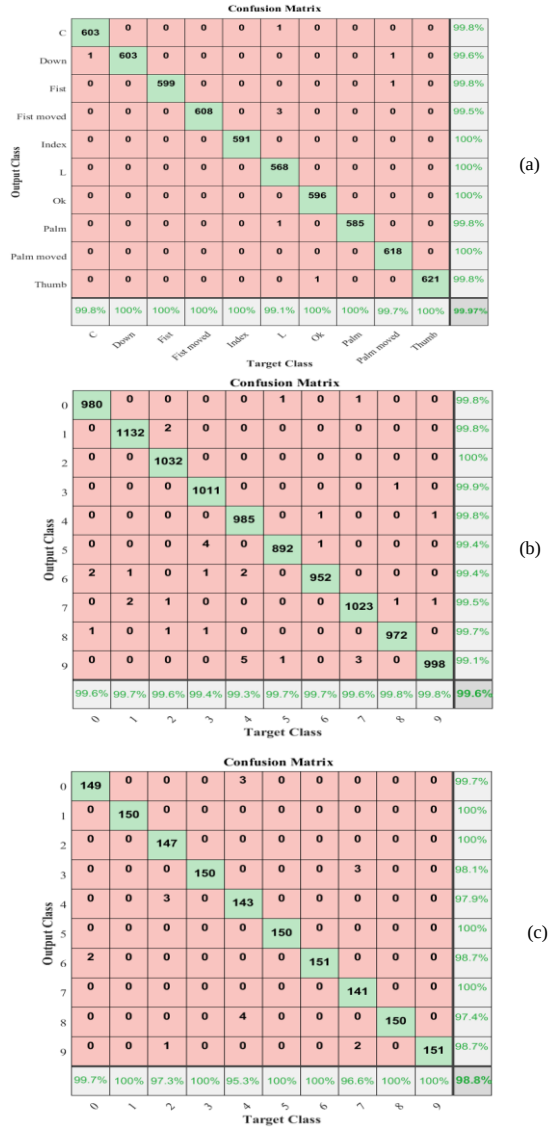


Fig. 6. Confusion matrix of the proposed model $VRFNet$ with (a) LeapGestRecog (b) MNIST digits and (c) ASL digits datasets for hand gesture recognition.

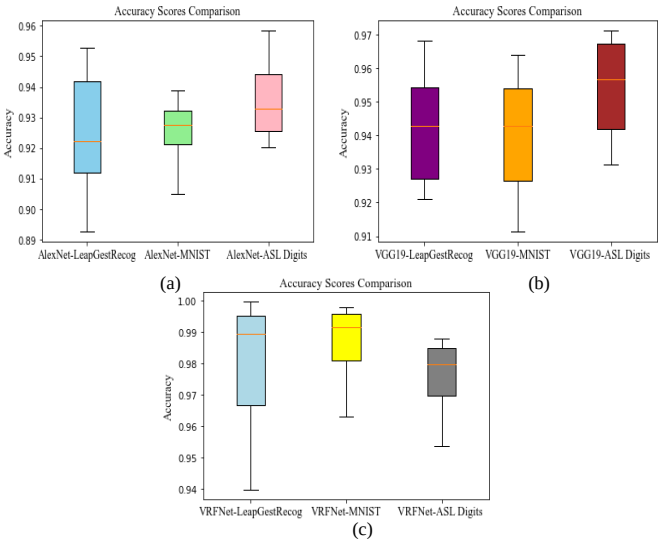


Fig. 7. Box plot of recognition accuracy of tested models (a) AlexNet, (b) VGG19, and (c) $VRFNet$ model for hand gesture recognition.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

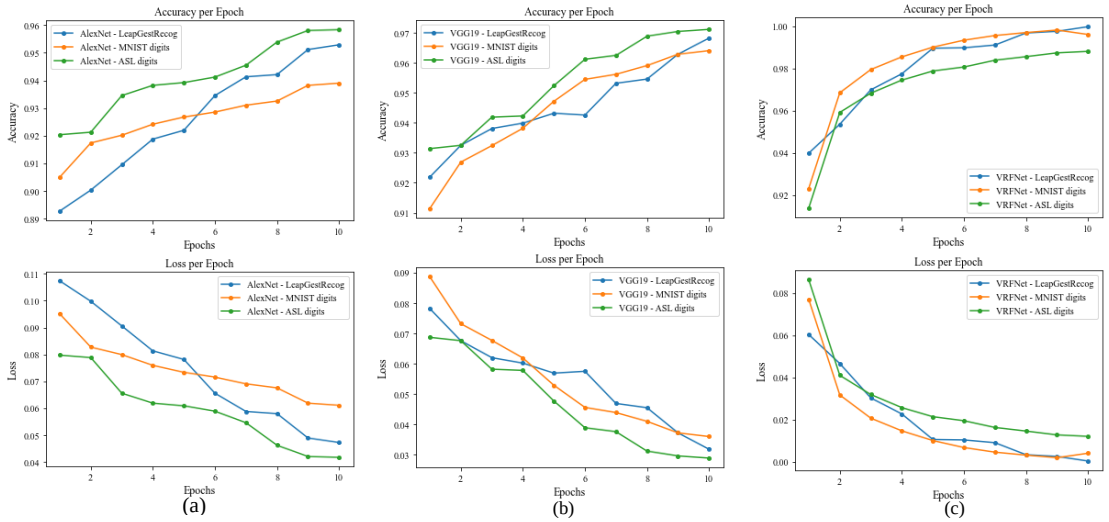


Fig. 8. Accuracy and loss curve of tested models (a) AlexNet (b) VGG19 (c) $VRFNet$ model with three datasets (LeapGestRecog, MNIST digits and ASL Digits) for hand gesture recognition.

TABLE II
COMPARISON OF THE PROPOSED METHODOLOGY WITH OTHER EXISTING METHODS USING SIMILAR DATASETS FOR HAND GESTURE RECOGNITION

Dataset	Authors	Technique	A_c (%)
LeapGest Recog	Mantecon et al. [35]	DSQP-based image descriptor + SVM	99.0
	Stellin et al. [38]	HandGCNN (modified CNN)	99.13
	Sharma et al. [41]	Enhanced CNN	99.7
	Proposed	***ASLiT + VRFNet***	**99.97**
MNIST Digits	Keerthi et al. [37]	CNN	99.2
	Shao et al. [42]	Hyper parameter optimization + Deep CNN	99.4
	Alqahtani et al. [43]	CNN	99.0
		DNN	99
		LeNet-5	91
		AlexNet	98
		MiniVGGNet	94
	Swain et al. [44]	ANN model employed with an Adagrad optimizer.	98.55
	Proposed	***ASLiT + VRFNet***	**99.6**
ASL Digits	Sahoo et al. [40]	CNN + dual stream dense residual fusion	96.14
	Alom et al. [45]	CNNSVM	98.2
	Kalam et al. [46]	CNN + residual learning	97.28
	Proposed	***ASLiT + VRFNet***	**98.8**

TABLE III
COMPARISON OF THE PROPOSED METHODOLOGY WITH OTHER EXISTING METHODS FOR HAND GESTURE RECOGNITION

Dataset	Authors	Technique	A_c (%)
NUS-II	Sahoo et al. [40]	DeReFNet	96.70
DHGD	Okano et al. [47]	ST-GCN	94.35
NUS-II	Sahoo et al. [48]	RBI-2RCNN	94.80
NUS-I	Bhaumik et al. [49]	ExtriDeNet	96.88
HGR-I			93.56
DHG-14	Li et al. [50]	self-attention-based graph convolutional network	96.31
DGH-28			94.05
FPHA			90.26
ASL-FS	Sahoo et al. [51]	DRCAM	93.44
ASL	Kankariya et al. [52]	Inception V3	98.54
LeapGest Recog	**Proposed**	***ASLiT + VRFNet***	**99.97**
MNIST Digits	**Proposed**	***ASLiT + VRFNet***	**99.6**
ASL Digits	**Proposed**	***ASLiT + VRFNet***	**98.8**

and 98.8% on the LeapGestRecog, MNIST digits and ASL digits datasets, respectively. From Table III, the exceptional performance of the $VRFNet$ architecture, leveraging the $ASLiT$ technique to improve recognition accuracy of hand gestures, is apparent. Additionally, the comparison of the proposed methodology employing AlexNet and $ASLiT$ with other wavelet transform based techniques CWT [53] and STFT [54] is presented in Table IV using LeapGestRecog dataset. The purpose of employing other wavelet transform based techniques is to assess the effect of $ASLiT$ on recognition accuracy compared to them. This comparison highlights the potential of $ASLiT$ in improving the accuracy of hand gesture recognition over CWT and STFT.

TABLE IV
COMPARISON OF THE PROPOSED *ASLiT* WITH OTHER WAVELET TRANSFORM-BASED TECHNIQUES FOR HAND GESTURE RECOGNITION USING LEAPGESTRECOG DATASET

Technique	Dataset	Classification Model	A_c (%)
CWT	LeapGestRecog	AlexNet	91.26
STFT		AlexNet	90.43
ASLiT		AlexNet	95.29

B. Ablation Study

To evaluate the impact of the proposed $VRFNet$ deep learning -based gesture classification model utilizing $ASLiT$ technique, an ablation study is conducted. Table V reports the progressive development of the $VRFNet$ model on different evaluation parameters. This analysis demonstrates the effect and contribution of the different models in enhancing the performance of the $VRFNet$ model. The study begins with model M1, that combines $ASLiT$ and AlexNet. This initial model, when compared to the Up-Down approach, lacks improved recognition accuracy of different classes. To enhance the M1, model M2 is introduced, incorporating a VGG19 in place of AlexNet. This upgraded model gives improved recognition accuracy but still model unable to distinguish between similar classes of some gestures.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

TABLE V
AN ABLATION STUDY OF THE PROPOSED *VRFNet* MODEL BASED ON DIFFERENT EVALUATION PARAMETERS FOR HAND GESTURE RECOGNITION

Model	Network	Dataset	A_c (%)	P_r (%)	R_l (%)	F_s (%)	T_r (hr.)	T_s (sec.)
M1	*ASLiT*+ AlexNet	LeapGestRecog	95.29	97.30	97.80	97.54	0.5	1.4
		MNIST digits	93.90	94.50	93.60	94.05	0.7	2.1
		ASL digits	95.84	96.90	96.20	96.54	0.3	1.2
M2	*ASLiT*+ VGG19	LeapGestRecog	96.83	98.50	98.35	98.42	0.6	1.1
		MNIST digits	96.41	98.23	98.14	98.18	0.8	2.3
		ASL digits	97.12	98.04	99.01	98.52	0.7	2.9
M3	*ASLiT*+ ReseNet50	LeapGestRecog	94.21	95.24	93.02	94.11	0.9	3.4
		MNIST digits	93.34	89.74	93.33	91.49	1.3	2.2
		ASL digits	95.73	92.11	89.74	90.90	0.8	1.3
M4	*ASLiT*+ AlexNet+ VGG19	LeapGestRecog	95.22	95.89	90.91	93.33	1.1	2.6
		MNIST digits	94.16	93.24	94.35	93.79	1.3	2.1
		ASL digits	93.46	93.29	92.56	92.92	0.8	1.4
M5	*ASLiT*+ AlexNet+ ResNet50	LeapGestRecog	94.43	94.63	94.29	94.45	1.2	2.2
		MNIST digits	93.14	94.38	93.65	94.01	1.3	1.8
		ASL digits	94.86	94.29	93.41	93.84	0.7	1.2
M6	*ASLiT*+ VGG19 + ResNet50	LeapGestRecog	**99.97**	**99.83**	**99.86**	**99.85**	**0.9**	**3.5**
		MNIST digits	**99.60**	**99.64**	**99.62**	**99.63**	**1.2**	**4.1**
		ASL digits	**98.80**	**99.05**	**98.89**	**98.96**	**0.5**	**2.7**

M1, M2, M3, M4, M5, M6 represents model number, T_r is training time in hours, T_s is testing time in seconds.

To achieve better performance, model M3, M4 and M5 models are introduces which incorporates *ASLiT* with ResNet50, fusion of VGG19 and AlexNet, fusion of AlexNet and ResNet50 respectively. These approaches provide inferior evaluation parameters with respect to M1 and M2. Furthermore, model M6 which incorporates *ASLiT* with the fusion of VGG19 and ResNet50 provides outperforming results. In terms of evaluation parameters, improvement is observed in accuracy, precision, recall and F1 score with less training and testing time. This further emphasizes the benefits of employing model M6 in facilitating comprehensive and effective multimodal analysis of hand gesture recognition system.

V. CONCLUSION AND FUTURE SCOPE

The paper presents an Adaptive Superlet Transform for images which provides high-resolution in localization of both time and frequency simultaneously. To the best of our knowledge this is the first attempt that incorporates *ASLiT* for images. *ASLiT* is further incorporated with *VRFNet* that is a fusion of VGG19 and ResNet50 for the classification of hand gestures. Each model has learned different high-level features or captured different patterns, and combining them can provide a more comprehensive representation of the data. Extensive experiments are conducted on LeapGestRecog, MNIST digits, ASL digits datasets to validate the efficiency and efficacy of the proposed framework *VRFNet*. The proposed model provided a recognition accuracy of 99.97%, 99.6% and 98.8% LeapGestRecog, MNIST digits and ASL digits datasets, respectively. To evaluate the proficiency of the proposed methodology employing *ASLiT* in contrast to other transform-based techniques, the model is also assessed on the spectrograms generated by CWT and STFT, which yield subpar results. An ablation study is also carried out to support the experimental analysis. The findings highlight the importance of the *VRFNet* architecture and the utility of the *ASLiT* technique in improving recognition accuracy compared with several state-of-the-art techniques.

Future research may explore the applicability of the proposed framework to other modalities, such as video-based or depth-based hand gesture recognition, expanding to larger and more diverse datasets, could be a fruitful area of research.

REFERENCES

[1] Md. A. Rahim, J. Shin, and Md. R. Islam, "Hand gesture recognition-based non-touch character writing system on a virtual keyboard," *Multimedia Tools and Applications*, vol. 79, no. 17–18, pp. 11813–11836, Jan. 2020, doi: 10.1007/s11042-019-08448-6.

[2] W. K. Wong, F. H. Juwono, and B. T. T. Khoo, "Multi-Features Capacitive Hand Gesture Recognition Sensor: A Machine Learning Approach," *IEEE Sensors Journal*, vol. 21, no. 6, pp. 8441–8450, Mar. 2021, doi: 10.1109/jsen.2021.3049273.

[3] P. S. Neethu, R. Suguna, and D. Sathish, "An efficient method for human hand gesture detection and recognition using deep learning convolutional neural networks," *Soft Computing*, vol. 24, no. 20, pp. 15239–15248, Mar. 2020, doi: 10.1007/s00500-020-04860-5.

[4] J. P. Vásconez, L. I. B. López, Á. L. V. Caraguay, P. J. Cruz, R. Álvarez, and M. E. Benalcázar, "A Hand Gesture Recognition System Using EMG and Reinforcement Learning: A Q-Learning Approach," *Lecture Notes in Computer Science*, pp. 580–591, 2021, doi: 10.1007/978-3-030-86380-7_47.

[5] A. Wadhawan and P. Kumar, "Deep learning-based sign language recognition system for static signs," *Neural Computing and Applications*, vol. 32, no. 12, pp. 7957–7968, Jan. 2020, doi: 10.1007/s00521-019-04691-y.

[6] P. Tsinganos, B. Cornelis, J. Cornelis, B. Jansen, and A. Skodras, "Hilbert sEMG data scanning for hand gesture recognition based on deep learning," *Neural Computing and Applications*, vol. 33, no. 7, pp. 2645–2666, Jul. 2020, doi: 10.1007/s00521-020-05128-7.

[7] M. A. Ozdemir, D. H. Kisa, O. Guren, and A. Akan, "Hand gesture classification using time–frequency images and transfer learning based on CNN," *Biomedical Signal Processing and Control*, vol. 77, p. 103787, Aug. 2022, doi: 10.1016/j.bspc.2022.103787.

[8] S. E. Agab and F. Z. Chelali, "New combined DT-CWT and HOG descriptor for static and dynamic hand gesture recognition," *Multimedia Tools and Applications*, vol. 82, no. 17, pp. 26379–26409, Jan. 2023, doi: 10.1007/s11042-023-14433-x.

[9] H. Qu, T. Li, R. Wang, J. Li, Z. Guan, and G. Chen, "Application of adaptive wavelet transform based multiple analytical mode decomposition for damage progression identification of Cable-Stayed bridge via shake table test," *Mechanical Systems and Signal Processing*, vol. 149, p. 107055, Feb. 2021, doi: 10.1016/j.ymssp.2020.107055.

[10] V. V. Moca, H. Bârzan, A. Nagy-Dăbâcan, and R. C. Mureșan, "Time-frequency super-resolution with superlets," *Nature Communications*, vol. 12, no. 1, Jan. 2021, doi: 10.1038/s41467-020-20539-9.

[11] H. Barzan, V. V. Moca, A.-M. Ichim, and R. C. Muresan, "Fractional Superlets," *2020 28th European Signal Processing Conference (EUSIPCO)*, Jan. 2021, doi: 10.23919/eusipco47968.2020.9287873.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

[12] G. Park, V. K. Chandrasegar, and J. Koh, "Accuracy Enhancement of Hand Gesture Recognition Using CNN," *IEEE Access*, vol. 11, pp. 26496–26501, 2023, doi: 10.1109/access.2023.3254537.

[13] M. M. Damaneh, F. Mohanna, and P. Jafari, "Static hand gesture recognition in sign language based on convolutional neural network with feature extraction method using ORB descriptor and Gabor filter," *Expert Systems with Applications*, vol. 211, p. 118559, Jan. 2023, doi: 10.1016/j.eswa.2022.118559.

[14] G. Bhaumik, M. Verma, M. C. Govil, and S. K. Vipparthi, "HyFiNet: Hybrid feature attention network for hand gesture recognition," *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 4863–4882, Jan. 2022, doi: 10.1007/s11042-021-11623-3.

[15] F. Duan, X. Ren, and Y. Yang, "A Gesture Recognition System Based on Time Domain Features and Linear Discriminant Analysis," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 13, no. 1, pp. 200–208, Mar. 2021, doi: 10.1109/tcds.2018.2884942.

[16] Y. Hu, Y. Wong, W. Wei, Y. Du, M. Kankanhalli, and W. Geng, "A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition," *PLOS ONE*, vol. 13, no. 10, p. e0206049, Oct. 2018, doi: 10.1371/journal.pone.0206049.

[17] U. Cote-Allard *et al.*, "Deep Learning for Electromyographic Hand Gesture Signal Classification Using Transfer Learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 760–771, Apr. 2019, doi: 10.1109/tnsre.2019.2896269.

[18] S. Pancholi and A. M. Joshi, "Improved Classification Scheme Using Fused Wavelet Packet Transform Based Features for Intelligent Myoelectric Prostheses," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 10, pp. 8517–8525, Oct. 2020, doi: 10.1109/tie.2019.2946536.

[19] S. Tam, M. Boukadoum, A. Campeau-Lecours, and B. Gosselin, "A Fully Embedded Adaptive Real-Time Hand Gesture Classifier Leveraging HD-sEMG and Deep Learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 2, pp. 232–243, Apr. 2020, doi: 10.1109/tbcas.2019.2955641.

[20] A. Tripathi, A. P. Prathosh, S. P. Muthukrishnan, and L. Kumar, "SurfMyoAiR: A Surface Electromyography-Based Framework for Airwriting Recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023, doi: 10.1109/tim.2023.3248084.

[21] Z. Yang *et al.*, "Dynamic Gesture Recognition Using Surface EMG Signals Based on Multi-Stream Residual Network," *Frontiers in Bioengineering and Biotechnology*, vol. 9, Oct. 2021, doi: 10.3389/fbioe.2021.779353.

[22] P. Balaji and M. Ranjan Prusty, "Multimodal fusion hierarchical self-attention network for dynamic hand gesture recognition," *Journal of Visual Communication and Image Representation*, vol. 98, p. 104019, Feb. 2024, doi: 10.1016/j.jvcir.2023.104019.

[23] R. Rastgoo, K. Kiani, S. Escalera, and M. Sabokrou, "Multi-modal zero-shot dynamic hand gesture recognition," *Expert Systems with Applications*, vol. 247, p. 123349, Aug. 2024, doi: 10.1016/j.eswa.2024.123349.

[24] P. M. Tripathi, A. Kumar, M. Kumar, and R. Komaragiri, "Multilevel Classification and Detection of Cardiac Arrhythmias with High-Resolution Superlet Transform and Deep Convolution Neural Network," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022, doi: 10.1109/tim.2022.3186355.

[25] Pankaj, A. Kumar, M. Kumar, and R. Komaragiri, "STSR: Spectro-Temporal Super-Resolution Analysis of a Reference Signal Less Photoplethysmogram for Heart Rate Estimation During Physical Activity," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022, doi: 10.1109/tim.2022.3192831.

[26] K. Bhatt, N. Jayanthi, and M. Kumar, "High-resolution superlet transform based techniques for Parkinson's disease detection using speech signal," *Applied Acoustics*, vol. 214, p. 109657, Nov. 2023, doi: 10.1016/j.apacoust.2023.109657.

[27] Pankaj, A. Kumar, R. Komaragiri, and M. Kumar, "Blood pressure estimation and classification using a reference signal-less photoplethysmography signal: a deep learning framework," *Physical and Engineering Sciences in Medicine*, vol. 46, no. 4, pp. 1589–1605, Sep. 2023, doi: 10.1007/s13246-023-01322-8.

[28] Pankaj, A. Kumar, M. Kumar, and R. Komaragiri, "Optimized deep neural network models for blood pressure classification using Fourier analysis-based time–frequency spectrogram of photoplethysmography signal," *Biomedical Engineering Letters*, vol. 13, no. 4, pp. 739–750, Jun. 2023, doi: 10.1007/s13534-023-00296-6.

[29] S. Singhal and M. Kumar, "GSMD-SRST: Group Sparse Mode Decomposition and Superlet-Transform-Based Technique for Multilevel Classification of Cardiac Arrhythmia," *IEEE Sensors Journal*, vol. 24, no. 6, pp. 8160–8169, Mar. 2024, doi: 10.1109/jsen.2024.3354113.

[30] P. M. Tripathi, A. Kumar, M. Kumar, and R. S. Komaragiri, "Automatic seizure detection and classification using super-resolution superlet transform and deep neural network -A preprocessing-less method," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107680, Oct. 2023, doi: 10.1016/j.cmpb.2023.107680.

[31] P. Srikanth and C. Koley, "An intelligent algorithm for autorecognition of power system faults using superlets," *Sustainable Energy, Grids and Networks*, vol. 26, p. 100450, Jun. 2021, doi: 10.1016/j.segan.2021.100450.

[32] S. Mitra and C. Koley, "Early and Intelligent Bearing Fault Detection Using Adaptive Superlets," *IEEE Sensors Journal*, vol. 23, no. 7, pp. 7992–8000, Apr. 2023, doi: 10.1109/jsen.2023.3245186.

[33] C. Yi *et al.*, "High-order Synchrosqueezing Superlets Transform and its application to mechanical fault diagnosis," *Applied Acoustics*, vol. 204, p. 109226, Mar. 2023, doi: 10.1016/j.apacoust.2023.109226.

[34] H. A. Agboola and J. E. Zaccheus, "Wavelet image scattering based glaucoma detection," *BMC Biomedical Engineering*, vol. 5, no. 1, Mar. 2023, doi: 10.1186/s42490-023-00067-5.

[35] T. Mantecón, C. R. del-Blanco, F. Jaureguizar, and N. García, "Hand Gesture Recognition Using Infrared Imagery Provided by Leap Motion Controller," *Advanced Concepts for Intelligent Vision Systems*, pp. 47–57, 2016, doi: 10.1007/978-3-319-48680-2_5.

[36] R. Stellin, P. Rukmani, L. J. Anbarasi, and S. Narayanan, "HandGCNN model for gesture recognition-based voice assistance," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 42353–42369, Aug. 2022, doi: 10.1007/s11042-022-13497-5.

[37] E. Rani. G, Sakthimohan. M, A. Reddy. G, Selvalakshmi. D, T. Keerthi, and R. Sekar. R, "MNIST Handwritten Digit Recognition using Machine Learning," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Apr. 2022, doi: 10.1109/icacite53722.2022.9823806.

[38] S. S. Ahmed, Z. Mehmood, I. A. Awan, and R. M. Yousaf, "A Novel Technique for Handwritten Digit Recognition Using Deep Learning," *Journal of Sensors*, vol. 2023, pp. 1–15, Jan. 2023, doi: 10.1155/2023/2753941.

[39] Barczak, Andre LC, et al. "A new 2D static hand gesture colour image dataset for ASL gestures." (2011).

[40] J. P. Sahoo, S. P. Sahoo, S. Ari, and S. K. Patra, "DeReFNet: Dual-stream Dense Residual Fusion Network for static hand gesture recognition," *Displays*, vol. 77, p. 102388, Apr. 2023, doi: 10.1016/j.displa.2023.102388.

[41] H. Kumar Sharma, P. Ahlawat, P. Kumar, N. Parashar, R. Soni, and Y. Manchanda, "DEEP LEARNING BASED ACCURATE HAND GESTURE RECOGNITION USING ENHANCED CNN MODEL." *Proceedings of Second International Conference on Computing*. 2021.

[42] H. Shao, E. Ma, M. Zhu, X. Deng, and S. Zhai, "MNIST Handwritten Digit Classification Based on Convolutional Neural Network with Hyperparameter Optimization," *Intelligent Automation & Soft Computing*, vol. 36, no. 3, pp. 3595–3606, 2023, doi: 10.32604/iasc.2023.036323.

[43] A. S. Alqahtani, A. N. Madheswari, A. Mubarakali, and P. Parthasarathy, "Secure communication and implementation of handwritten digit recognition using deep neural network," *Optical and Quantum Electronics*, vol. 55, no. 1, Nov. 2022, doi: 10.1007/s11082-022-04290-7.

[44] D. Swain, B. Parmar, H. Shah, A. Gandhi, B. Acharya, and Y.-C. Hu, "Enhanced handwritten digit recognition using optimally selected optimizer for an ANN," *Multimedia Tools and Applications*, vol. 82, no. 28, pp. 44021–44036, Apr. 2023, doi: 10.1007/s11042-023-15402-0.

[45] Md. S. Alom, Md. J. Hasan, and Md. F. Wahid, "Digit Recognition in Sign Language Based on Convolutional Neural Network and Support Vector Machine," *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dec. 2019, doi: 10.1109/sti47673.2019.9067999.

[46] Md. A. Kalam, Md. N. I. Mondal, and B. Ahmed, "Rotation Independent Digit Recognition in Sign Language," *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, Feb. 2019, doi: 10.1109/ecace.2019.8679172.

[47] M. Okano, J.-Q. Liu, T. Tateyama, and Y.-W. Chen, "DHGD: Dynamic Hand Gesture Dataset for Skeleton-Based Gesture Recognition and Baseline Evaluations," *2024 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2024, doi: 10.1109/icce59016.2024.10444226.

[48] J. P. Sahoo, S. P. Sahoo, S. Ari, and S. K. Patra, "RBI-2RCNN: Residual Block Intensity Feature using a Two-stage Residual Convolutional Neural Network for Static Hand Gesture Recognition," *Signal, Image and Video Processing*, vol. 16, no. 8, pp. 2019–2027, Feb. 2022, doi: 10.1007/s11760-022-02163-w.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Sensors Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2024.3424855

[49] G. Bhaumik, M. Verma, M. C. Govil, and S. K. Vipparthi, "ExtriDeNet: an intensive feature extrication deep network for hand gesture recognition," *The Visual Computer*, vol. 38, no. 11, pp. 3853–3866, Jul. 2021, doi: 10.1007/s00371-021-02225-z.

[50] C. Li, S. Li, Y. Gao, X. Zhang, and W. Li, "A Two-Stream Neural Network for Pose-Based Hand Gesture Recognition," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 4, pp. 1594–1603, Dec. 2022. Doi: 10.1109/tcds.2021.3126637.

[51] J. P. Sahoo, S. P. Sahoo, S. Ari, and S. K. Patra, "Hand Gesture Recognition Using Densely Connected Deep Residual Network and Channel Attention Module for Mobile Robot Control," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2023, doi: 10.1109/tim.2023.3246488.

[52] S. Kankariya, K. Thakre, U. Solanki, S. Mali, and A. Chunawale, "Sign Language Gestures Recognition using CNN and Inception v3," *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Mar. 2024, doi: 10.1109/esci59607.2024.10497401.

[53] J. E. W. Koh *et al.*, "Diagnosis of retinal health in digital fundus images using continuous wavelet transform (CWT) and entropies," *Computers in Biology and Medicine*, vol. 84, pp. 89–97, May 2017, doi: 10.1016/j.compbiomed.2017.03.008.

[54] K. Ciężar and M. Pochylski, "2D Short-Time Fourier Transform for local morphological analysis of meibomian gland images," *PLOS ONE*, vol. 17, no. 6, p. e0270473, Jun. 2022, doi: 10.1371/journal.pone.0270473.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on July 22,2024 at 10:27:23 UTC from IEEE Xplore. Restrictions apply.
© 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Check for updates

Received: 11 March 2024 | Revised: 9 May 2024 | Accepted: 13 June 2024

DOI: 10.1002/pa.2934

RESEARCH ARTICLE

WILEY

Workplace gratitude studies: Review and future agenda using bibliometric analysis

Naval Garg [1] | B. K. Punia [2]

[1]University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India

[2]Haryana School of Business, GJUS&T, Hisar, India

Correspondence
Naval Garg, University School of Management and Entrepreneurship, Delhi Technological University, Vivek Vihar, Phase-1, New Delhi, India.
Email: naval.garg@dtu.ac.in

The present study reviewed gratitude at work articles published in Scopus database using bibliometric analysis to examine growth in gratitude at work publications and citations, most cited publications, most influential authors, articles, publishers, universities and countries. The extracted data were analysed and visualised with the help of two analytical tools, that is, the Scopus analysis and VOSviewer (version 1.6.16). It identified 273 publications referring to gratitude at work in title, abstract, or keywords. The keyword co-occurrence highlighted four major clusters of workplace gratitude research: Psychological studies, Customer relationship management, Employees, and Ethics and spirituality.

KEYWORDS
bibliometric, trend analysis, Vos viewer, workplace gratitude

1 | INTRODUCTION

Modern workplaces are often characterised with a toxic environment that breeds egocentrism, deviance, incivility, harassments, office politics, glass ceiling, favouritism, and corruption (Fehr et al., 2017). The repercussions of such workplace toxicity could be seen at both levels: employee and organisational. At an individual level, it poorly affects employees' engagement, performance, self-esteem, psychological capital, work-life balance, physical and mental health, career progression, and stay in an organisational (Rasool et al., 2021). At the organisational level, it manifests as toxic culture which breeds toxic employees, toxic leaders, toxic processes and policies (Rasool et al., 2021). It creates a vicious circle wherein individual and organisational toxicity promotes each other. Luckily, organisations are trying out diverse means like promoting spirituality and gratitude to break this ominous cycle of toxicity. A few previous researchers have reported positive effects of gratitude at work. Gratitude improves life satisfaction (McCullough et al., 2002), reduces aggression (DeWall et al., 2012), motivates prosocial behaviour (Bartlett & DeSteno, 2006), supports work-family enrichment (Garg, 2020), creates positive relationships, social support, and workers' well-being (Di Fabio et al., 2017), enhances vitality and resilience (Garg et al., 2022a), promotes pro-social behaviour and organisational citizenship behaviour (Garg, 2017) and counters turnover and absenteeism (Garg et al., 2020). With the emergence of positive psychology, gratitude at work is increasingly advocated to promote workplace happiness, employee satisfaction, and humanity-based administration.

Although studies on gratitude have captured the attention of modern scientists in the last 34 years, exploration of gratitude at the workplace still remains a laggard (Fehr et al., 2017). There seems to be an interest in gratitude at work in the last couple of decades only. Nearly 55% of the organisational behaviour (OB) papers on gratitude have been published in the last 5 years (Locklear et al., 2022). There could be two possible reasons for ignoring gratitude at the workplace. Firstly, in the initial years, gratitude was a subject matter for philosophical and religious studies. The desirability of a secular and nonreligious workplace might have eluded the notion of grateful organisations from the mind of academicians, researchers and practitioners. Secondly, gratitude in work and non-work settings was not differentiated, thus, the findings of the studies conducted outside an organisation deem suitable and applicable on employees in professional settings too. But later on studies argued that one who is generally grateful in his/her personal life may not necessarily be so at work (Cain et al., 2019). Also, the diverse nature of challenges and aspiration at work offer different playing fields to gratitude. Given the benefits associated with practising gratitude at work and potential

 © 2024 John Wiley & Sons Ltd.

difference in experience of gratitude in work and non-work settings, it is imperative to have an in-depth understanding of various intricacies of gratitude at the workplace.

Accordingly, the present study attempts to review published manuscripts on gratitude at work using bibliometric analysis. Although there is a steep rise in the literature of gratitude at work research, there is no attempt to systematically identify the historical development and recent trends in workplace gratitude research. To the best of our knowledge and explorations, no previous researcher has reviewed and analysed the trends within the rapidly expanding domain of scientific knowledge of gratitude at work. There have been a few meta-analysis and systematic literature review articles on gratitude (see Boggiss et al., 2020; Davis et al., 2016), but bibliometric analysis are better than meta-analysis and systematic literature review studies in terms of its scope, coverage, utility, objectivity, and methodology (Donthu et al., 2021). As far as meta-analysis is concerned, it explains (1) the overall strength and direction of effects or relationships," and (2) "the across-study variance in the distribution of effect-size estimates and the factors that explain this variance" (Aguinis et al., 2011, p. 310). They are restricted to the examination of direction and strength of effects and relationships among two or more variables (Donthu et al., 2021). Further, systematic literature review uses systematic methodologies to acquire, arrange, and assess available manuscripts and articles. They are better suited for confined or niche research areas like domain-based reviews, methods-based reviews, or interventions-based reviews.

In contrast, bibliometric analysis studies are not restricted to any specific sub-theme, intervention, practice, or association between two variables. It is a scientifically developed mechanism introduced by Pritchard (1969) to logically examine historical developments, emerging trends, active research areas, and future directions in a particular field of interest. It has been defined as "the application of mathematical and statistical methods to articles and other means of communication" (Pritchard, 1969, p. 349). The bibliometric analysis is often advocated for its objectivity and ability to handle large data and files using mathematical and statistical methods (Donthu et al., 2021). It offers bibliometric and intellectual structure of available literature of the field of study through examining the social and structural relationships between different research constituents (e.g., authors, countries, institutions, and topics) (Donthu et al., 2021). The bibliometric analysis indicates rate, state, and direction of development of knowledge in a particular area. The present study offers an extensive understanding of publications, citations, journals, countries and authors in the domain of gratitude at work using bibliometric analysis. It highlights prominent/most contributing authors, institutions, and countries. It also captures whether the area of investigation is dominated by a few scholars, universities, or regions. It explains how gratitude at work research has developed and disseminated over the last few decades. It evaluates historical developments, emerging trends, and active research areas in the field of gratitude at work. The bibliometric analysis clearly delineates "what has already been examined," and "what has not been examined so far;" thus, it makes a researcher aware about development till date and the research gaps. Thereby, it acts as a guiding light for future researchers. "It enables us to unpack the evolutionary nuances of a specific field, while shedding light on the emerging areas in that field" (Donthu et al., 2021, p. 285). Accordingly, this study performs a bibliometric analysis of gratitude at work research to answer the following research questions. The present study utilised the stakeholders' perspective to explore the gratuitous behaviour of different stakeholders such as employees, customers, and suppliers. It is important to note that gratitude is contagion and one's gratuitous behaviour influences others to reciprocate in the similar fashion (Garg, 2024). Organisations put effort to elicit gratitude among its different players like leaders, employees, and consumers; thus, a study on gratitude at work needs to capture gratitude among stakeholders of a firm.

RQ1. What are the current publication and citation trends of gratitude at work research?

RQ2. Which are the most influential authors, articles, journals, publishers, and countries publishing extensively in the domain of gratitude at work?

RQ3. What are the different research areas that carry out research in the gratitude at work domain?

RQ4. Which themes of gratitude at work contemporary researchers are exploring, and also which topics have scope for future research?

2 | METHOD

2.1 | Data source, search strategy, and data retrieval

The Scopus database was used to extract data for bibliometric analysis. The Scopus is the largest database of peer-reviewed scientific literature including journals, books, and conference proceedings. Other databases like Google scholar, Pubmed, PsycInfo, and Web of Science (WoS) are also available but the Scopus database is used for its greater relevance in the present context. Google scholar does not provide a dataset in such a form suited for bibliometric analysis. The PubMed dataset predominantly comprises research in life sciences and biomedical topics. And the PsycInfo database consists of works in the field of psychology. However, studies on gratitude at work are distributed in psychology, marketing, human resource, economics, medical, etc. Also, Scopus is the largest database and it provides information about the most prominent authors, countries, affiliations, journals, and publication years both in tabular and graphic forms. Therefore, the Scopus database is preferred over all other available databases. The phrases "gratitude and work," "gratitude and workplace," "gratitude and organisations," "gratitude and company," "gratitude and worker," "gratitude and employee," "gratitude and profession," "gratitude and customer," "gratitude and consumer," and

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

"gratitude and client," were searched on Scopus database in the title, abstract, or keywords of the publications as on 28th February 2023. Only documents of the article and review types were considered, and other document types like books, chapters, and letters were discarded. Articles published in journals are subjected to stringent review process conducted by editors, associate editors, and peer reviewers, which ensures the quality of the publication leading to greater acceptability of the results among academicians. On contrary, books and book chapters do not follow rigorous review process; thus they are excluded from the present study. Further, selection was restricted to "business, management, and accounting" category publications in English language only. Gratitude studies are widely spread across various disciplines such as psychology, sociology, ethics and spirituality, medicine, psychiatry, nursing, and philosophy. Since the present study intends to evaluate the gratitude at work; thus, the scope of the study was restricted to "business, management, and accounting." The search string resulted in 347 studies, which were manually screened through reading their abstract. Ultimately, 273 publications were selected from further analysis and deliberations.

2.2 | Data analysis and visualisation

The extracted data were analysed and visualised with the help of two analytical tools, that is, the Scopus analysis and VOSviewer (version 1.6.16). The Scopus database provides information about the most prominent authors, countries, affiliations, journals, and publication years. The VOSviewer provides visualisation facilities for country collaboration, co-authorship, and keyword co-occurrence. In this study, the number of publications and citations over years, most cited publications, and most influential authors, articles, publishers, universities, and countries are elicited as a part of performance analysis. Also, major research areas of gratitude research are also explored. Furthermore, collaboration among countries, keyword co-occurrence, and recent trends analysis are employed through science mapping. Donthu et al. (2021) recommended that performance analysis and co-citation analysis reveal the knowledge foundation (past knowledge) and keyword co-occurrence analysis reveals the knowledge production (present trend of knowledge development). Accordingly, this study employs both methods to deep dive into the ocean of knowledge of gratitude research.

3 | RESULTS

3.1 | Number of publications over years

Scopus database identified 273 publications referring to given keywords. It comprised 263 articles and 10 reviews. In the most recent article, published in *Journal of Service Marketing*, Qi et al. (2023) reported that the organisations' service climate leads to employee gratitude, which ultimately results in customer oriented behaviour by employees. In another recent article, Zhao et al. (2023) suggested a positive relationship between gratitude and employee well-being and organisational commitment. Although the first research article appeared in 1982, there has been an acute dearth of research until 2010. The scientific exploration of gratitude at the workplace gained momentum after 2010. The year 2010 witnessed only 1 publication and it rose to 44 publications by 2022. There have been 13 publications in the first 2 months of 2023. The last 12 years witnessed an exponential increase in publication with 229 publications and the last 4 years (2019 to Feb 2023) observed 129 publications. These trends clearly suggest a growing interest of academics and practitioners in developing gratuitous workplaces and experiences. Figure 1 provides the year-wise growth of publication referring to gratitude at the workplace

3.2 | Most cited publications

Table 1 reports the top 10 cited publications in the domain of gratitude at work. The most cited publication is "*The role of customer gratitude in relationship marketing*." It is published in the *Journal of Marketing* and has been cited 478 times since 2009. It also has the highest citation per year score of 31.87 citations per year. The study offers an important insight into how gratitude helps relationship marketing to achieve their intended performance outcomes. It reports that gratitude mediates the relationship between relationship marketing and purchase intentions, sales growth, and share of wallet. The second top cited publication is Morales's (2005) "*Giving firms an 'E' for effort: Consumer responses to high-effort firms*" published in the *Journal of Consumer Research* with 317 citations so far. The study reports that a firms' extra efforts in manufacturing, displaying, selling a product is rewarded by consumers' increasing their willingness to pay, store choice, and overall evaluations. Based on attribution theory, the findings of the study suggest that the proposed relationship is mediated by customers' feeling of gratitude. It is interesting to note that eight out of the ten top cited publications are in the domain of customer relationship management. One of the top cited articles beyond the scope of customer relationship is Fehr et al.'s (2017) "*The grateful workplace: A multilevel model of gratitude in organizations*" with 141 citations (20.14 citations per year). This article highlighted the scarcity of empirical studies on gratitude at the workplace and proposed a multilevel gratitude model for workplaces. It proposed a model consisting of episodic gratitude (at the event level), persistent gratitude (at the individual level), and collective gratitude (at the organisational level).

3.3 | Most prolific journal

There are 158 journals with at least one publication on gratitude at the workplace. Also, there are only seven journals with five or more publications and only 14 journals with four or more publications. The top three journals (in terms of number of publications) are *Journal of Business Research* (number of publications = 16, citation = 463, and

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

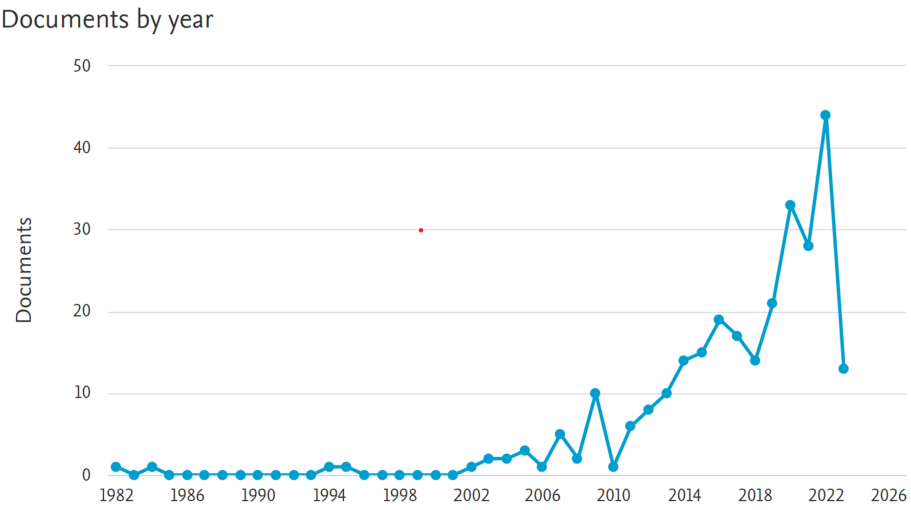


FIGURE 1 Year-wise growth in publication of gratitude at work research. *Source*: Delineated from Scopus database.

TABLE 1 Top 10 most cited publications on gratitude at work.

S. No	Author	Title of publication	Journal	Citations	Citation per year
1	Palmatier et al. (2009)	The role of customer gratitude in relationship marketing	Journal of Marketing	478	31.87
2	Morales (2005)	Giving firms an "E" for effort: Consumer responses to high-effort firms	Journal of Consumer Research	317	16.68
3	Soscia (2007)	Gratitude, delight, or guilt: The role of consumers' emotions in predicting post consumption behaviours	Psychology and Marketing	216	12.71
4	Romani et al. (2013)	Explaining Consumer Reactions to Corporate Social Responsibility: The Role of Gratitude and Altruistic Values	Journal of Business Ethics	174	15.82
5	Xie et al. (2015)	The role of moral emotions and individual differences in consumer responses to corporate green and non-green actions	Journal of the Academy of Marketing Science	159	17.67
6	Fehr et al. (2017)	The grateful workplace: A multilevel model of gratitude in organisations	Academy of Management Review	141	20.14
7	Wetzel et al. (2014)	Gratitude versus entitlement: A dual process model of the profitability implications of customer prioritisation	Journal of Marketing	134	13.40
8	Simon and Tossan (2018)	Does brand-consumer social sharing matter? A relational framework of customer engagement to brand-hosted social media	Journal of Business Research	133	22.17
9	Van dierendonck and Patterson (2015)	Compassionate Love as a Cornerstone of Servant Leadership: An Integration of Previous Theorising and Research	Journal of Business Ethics	128	14.22
10	Steinhoff and Palmatier (2014)	Understanding loyalty program effectiveness: managing target and bystander effects	Journal of the Academy of Marketing Science	122	15.25

Source: Scopus database.

citation per year = 28.94), *Journal of Business Ethics* (number of publications = 10, citation = 545, and citation per year = 23.44), and *Journal of Service Marketing* (number of publications = 9, citation = 211, and citation per year = 23.44). Also, five out of top the 10 journals are directly related to marketing research. There are a few journals related to human resource and behaviour too. These journals are *Personnel Review*, *Journal of Organisational Behaviour*, and *Journal of Management, Spirituality, and Religion*. *Journal of Marketing* has highest number of citations (644) and citation per year (128.80). Table 2 reports 14 journals with four or more publications. Also, Figure 2 represents journal-wise trends in gratitude at work research publications, which suggests different clusters of journals. The earliest cluster shows publication in *Journal of Business Ethics* and *Journal of Management, Spirituality, and Religion*. Then a few clusters indicate prominence of marketing journals like *Journal of Service Marketing*, *Journal of Retailing and Consumer Services* and *Journal of Marketing*. And, the most recent cluster highlights publications in non-marketing journals like *Journal of Business Research*, *Personnel Review*, and *Journal of Organisational Behaviour*. These findings reveal a clear prominence of gratitude in marketing-related journals. However, recent trends suggest increasing publications in the domain of human resource and organisational behaviour. A few publications in 2023 reveal linkage between gratitude and employees' ability to take charge (Ren et al., 2023), workers' turnover intentions (Garg et al., 2022b),

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Journals with 4 or more publications on gratitude at work.

S. No	Title of the journal	No. of documents	Citations	Citation per document
1	Journal of Business Research	16	463	28.94
2	Journal of Business Ethics	10	545	54.50
3	Journal of Services Marketing	9	211	23.44
4	European Journal of Marketing	8	114	14.25
5	Journal of Retailing And Consumer Services	7	175	25.00
6	Journal of the Academy of Marketing Science	6	457	76.17
7	Journal of Marketing	5	644	128.80
8	International Journal of Contemporary Hospitality management	4	111	27.75
9	International Journal of Hospitality Management	4	106	26.50
10	Journal of Management Spirituality And Religion	4	61	15.25
11	Journal of Marketing Communications	4	60	15.00
12	Journal of Organisational Behaviour	4	113	28.25
13	Personnel Review	4	8	2.00
14	Psychology And Marketing	4	236	59.00

Source: Scopus database.

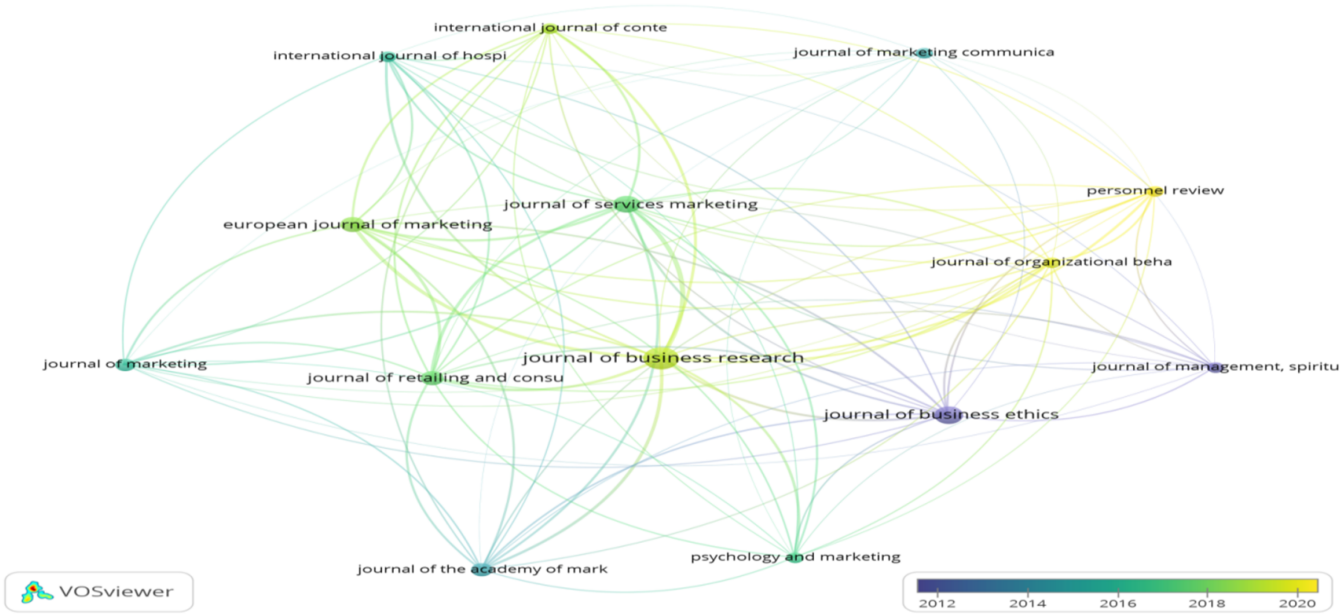


FIGURE 2 Recent journal-wise trends in gratitude at work publications. *Source*: Delineated from VosViewer.

workplace toxicity, employee well-being and organisational commitment (Zhao et al., 2023). It suggests gratitude practices are increasingly been used to solve organisational challenges in modern workplaces.

3.4 | Most prolific authors

Table 3 illustrates the list of authors with four or more publications on gratitude at work, along with their affiliation and country. There are three authors—Richard P. Bagozzi (University of Michigan, USA), Dora E. Bock (Auburn University, USA), and Judith Anne Garretson Folse (Louisiana State University, USA)—with a maximum eight publications on gratitude at work. It is interesting to note that five out of the top 10 authors are from the USA. Another notable contributor is Naval Garg (Delhi Technological University) with six publications. The author proposed gratitude as a construct of workplace spirituality and suggested institutionalisation of gratitude practices at the workplace (Saxena et al., 2020). In other papers, the authors reported positive association between gratitude and workplace happiness (Garg et al., 2022a), workplace toxicity (Sarkar et al., 2023), and technostress (Garg et al., 2022b). Two other top rated authors from Italy—Silvia Grappi and Simona Romani—have extensively examined the role of gratitude in customer relationship management.

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Authors with 4 or more publications on gratitude at work.

S. No	Author	Institution	Country	No. of publication	Citation	Citation per document
1	Richard P. Bagozzi	University of Michigen	USA	8	546	68.25
2	Dora E. Bock	Auburn University, Auburn, Alabama	USA	8	160	20.00
3	Judith Anne Garretson Folse	Louisiana State University, Baton Rouge	USA	8	281	35.13
4	Naval Garg	Delhi Technological university	India	6	23	3.83
5	Silvia Grappi	University of Modena and Reggio Emilia	Italy	5	287	57.40
6	Simona Romani	Luiss University	Italy	5	287	57.40
7	Felix Septianto	University of Queensland	Australia	5	114	22.80
8	Natalia Kolyesnikova	Texas Tech University	USA	4	117	29.25
9	Stephanie M. Mangus	Baylor University	USA	4	95	23.75
10	Gary Mortimer	QUT University	Australia	4	97	24.25
11	Françoise Simon	University of Haute-Alsace	France	4	207	51.75
12	Seon-Young Kim	National University Medical School	Republic of Korea	4	85	21.25

Source: Scopus database.

TABLE 4 Top 20 countries (number of documents).

S. No	Country/territory	No. of publications	Citations	Citations per document
1	USA	124	3972	32.03
2	India	26	110	4.23
3	Australia	22	436	19.82
4	China	22	158	7.18
5	United Kingdom	20	374	18.70
6	Germany	11	545	49.55
7	Taiwan	10	107	10.70
8	Italy	9	533	59.22
9	France	8	354	44.25
10	Hong Kong	8	126	15.75
11	Malaysia	8	29	3.63
12	New Zealand	7	172	24.57
13	South Korea	6	65	10.83
14	Spain	6	68	11.33
15	Canada	5	45	9.00
16	Pakistan	5	50	10.00
17	Netherlands	4	291	72.75
18	Norway	4	311	77.75
19	Portugal	4	2	0.50
20	Turkey	4	47	11.75

Source: Scopus database.

3.5 | Most prolific countries

Table 4 reports the top 20 countries in terms of the number of research publications in the domain of gratitude at work. The USA tops the list with 124 publications, 3972 citations, and 32.023 citations per year. A few prominent authors of the USA are Richard P. Bagozzi, Dora E. Bock, Judith Anne Garretson Folse, Natalia Kolyesnikova, and Stephanie M. Mangus. The next top ranked country is India (26 publications, 110 citations, and 4.23 citations per year), followed by Australia and China with a similar number of publications (22 publications). Australia has a relatively higher number of citations of 436 (citation per year = 19.82) than China's 158 (citation per

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

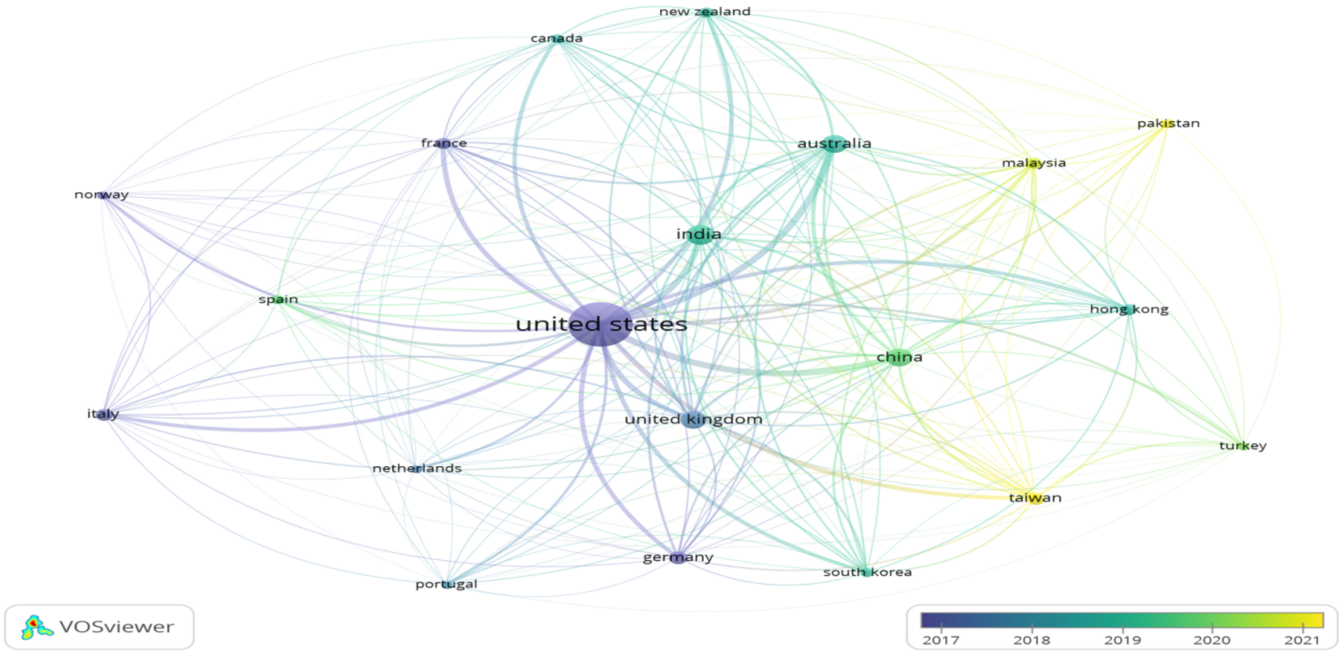


FIGURE 3 Recent country/territory-wise trends in gratitude at work publications. *Source*: Delineated by Vosviewer.

year = 7.18). Among these top 20 countries, the USA also has the highest number of citations (total number of citations = 3972), followed by Germany (total number of citations = 545) and Italy (total number of citations = 533). Norway (citations per document = 77.75), Netherlands (citations per document = 72.75), and Italy (citations per document = 59.22) are the three top most countries in terms of citation per documents. Interestingly, despite being in the top three countries (in terms of number of documents), India and China have relatively low citations per document with 4.23 and 7.18 citations per document. Figure 3 represents country/territory-wise trends in gratitude at work research work. It reports that earlier research was confined to the USA and European countries like UK, France, Germany, Norway, Italy, and Netherlands (highlighted in purple colour). Afterwards, scientific explorations of gratitude at work started in Australia, New Zealand and Asian countries like India, China, Hongkong, South Korea, and Turkey (highlighted in green colour). And recently, there is a growing interest in examining workplace gratitude in smaller countries like Pakistan, Malaysia, and Taiwan (highlighted in yellow colour). The figure suggests that gratitude at work research has spread over North America, Europe, Asia, and Australia continents; however South America and Africa are yet to contribute significantly in the domain of gratitude at the workplace research.

3.6 | Keywords co-occurrence

VosViewer was used to visualise keywords co-occurrence in the publications on gratitude at work. Figure 4 reports 10 keyword clusters, however four major clusters (with more than five items) are elaborated and considered for discussion. Further, Figure 5 reports most recent clusters or areas of workplace gratitude studies.

3.6.1 | Cluster-1: Gratitude and workplace well-being (red colour)

This cluster comprises 11 keywords like positive psychology, mindfulness, emotion, culture, psychological capital, and affect. A close examination of keywords hints toward two sub-themes. First, just like gratitude literature, gratitude at work has also been examined both as emotion and trait. Several publications including Schnall and Cannon (2012), and Liao et al. (2022) examined organisational benefits of gratitude as emotion. Other authors including Li et al. (2022), Zhoa et al. (2023), and Guzzo et al. (2020) explored trait gratitude at the workplace. Second, the cluster seems to capture the relationship between gratitude at work and workplace wellbeing. In a recent study, Sarkar et al. (2023) reported that employees' gratitude helps them counter workplace toxicity. They proposed that the negative relationship between gratitude and workplace toxicity is mediated by psychological capital. Based on the broaden-and-build theory, the authors stated that gratitude develops psychological capital (hope, resilience, self-efficacy, and optimism) among employees, which make them more competent to deal with workplace toxicity. Other psychological constructs like hope, loyalty, stress, burnout, trust, engagement, emotions, well-being, contentment, commitment, helping attitude, and happiness have also been examined in the light of gratitude. For instance, Zhoa et al. (2023) stated that gratitude positively influences individual (subjective well-being) and organisational (volunteer retention) outcomes through the mediating effects of prosocial motivation and

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

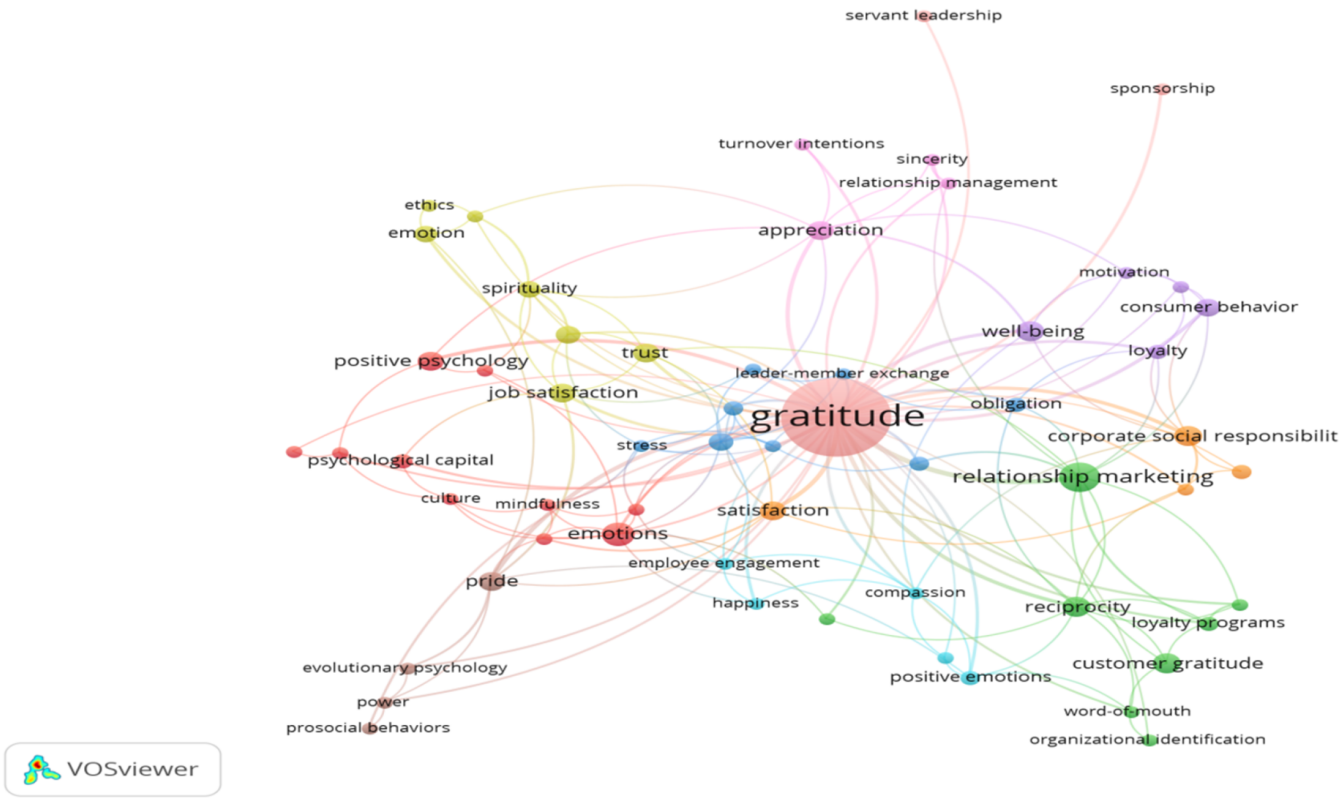


FIGURE 4 Keywords co-occurrence clusters in gratitude at work literature. *Source*: Delineated by Vosviewer.

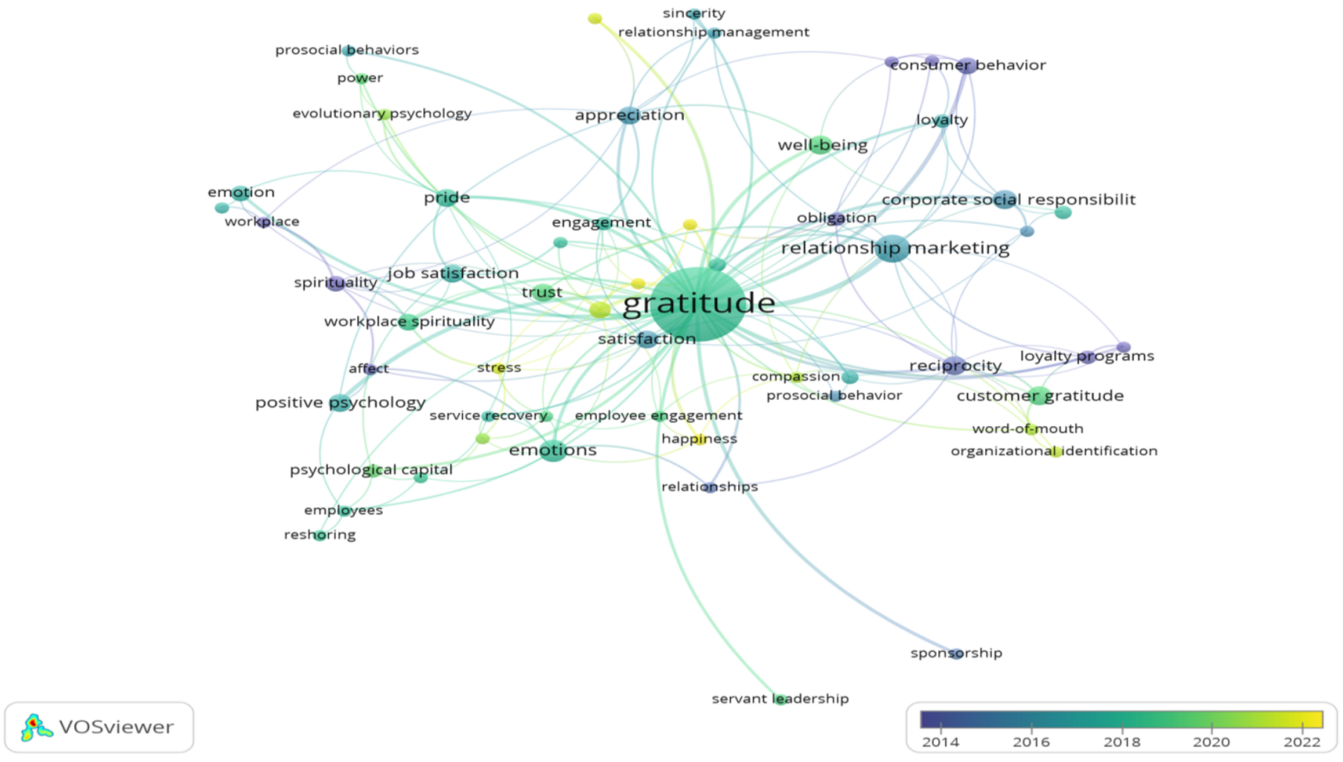


FIGURE 5 Recent clusters in gratitude at work literature. *Source*: Delineated by Vosviewer.

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

volunteer engagement. Similarly, Li et al. (2022) reported that gratitude increases people's well-being through connecting them to nature. The authors argued that gratitude leads to positive emotion of transcendence, which results in deep oneness with nature leading to physical and mental wellbeing. Besides, Garg et al. (2022c) suggested that gratitude enhances employees' workplace happiness through augmenting their social and psychological capital. Ali et al. (2022) also concluded that gratitude and mindfulness are pivotal for one's well-being. They elaborated that gratitude increases hopefulness and optimism leading to enhanced levels of wellbeing. In a nutshell, gratitude seems to influence workplace wellbeing in two ways. First, it augments different individual psychological constructs such as self-efficacy, resilience, hope, and optimism that results in workplace well-being. Second, gratuitous behaviour leads to better social relations and affiliations with other stakeholders like colleagues, supervisors, subordinates, and consumers, which results in accumulation of social capital essential for workplace wellbeing.

3.6.2 | Cluster-2: Gratitude and customer relationship management (green colour)

The second cluster has eight keywords including relationship, relationship marketing, reciprocity, customer gratitude, customer relationship management, and loyalty program. It focuses primarily on customer gratitude and its implication on customer relationship management. Two types of studies were reported. The first type of studies examined the direct effect of customer gratitude and the second type of studies explored the mediating or moderating role of customer gratitude. The studies reported linkage between customer gratitude and cooperation and helping other consumers (Chou et al., 2023), marketing investment (preferential treatment and interpersonal communication investments) (Mittal & Maity, 2022), territorial behaviour of consumers (Chou et al., 2023), customer loyalty (Mittal & Maity, 2022), relationship quality between salesperson and customer (Mangus et al., 2022), consumer intentions to spread electronic word-of-mouth (Septianto et al., 2020), individuals' sustainable product choice (Liang & Guo, 2021), and customer trust (Huang, 2015). Customer gratitude is reported to mediate the association between functional value and pay for price premium (Liu et al., 2022), brand likeability and brand love (Suetrong & Pires, 2021), consumer responses and marketing communications (Bridger & Wood, 2017), corporate philanthropy and consumer loyalty (Park et al., 2016), and human resource management attributions and employees general work stress (Hameed & Khwaja, 2022).

3.6.3 | Cluster-3: Gratitude at work and individual-level outcomes (blue colour)

The third cluster consists of eight keywords including employee performance, leader-member exchange, obligation, stress, engagement, and organisational commitment. This cluster seems to highlight various gratitude-driven individual outcomes in professional settings. While, the first cluster enumerates studies depicting association between gratitude and workplace wellbeing, this cluster encompasses studies highlighting other employees-related outcomes such as job satisfaction, team building, vitality, and proactive workplace behaviours. Again, two sub-themes were reported. In the first sub-theme, direct relationships were established between gratitude and employee-related variables. Various studies highlighted positive linkage between job satisfaction (Effendi et al., 2021), organisational commitment (Zhoa et al., 2023), take charge (Ren et al., 2023), work-to-family conflict (Zhan et al., 2022), employee performance during crisis (Ni et al., 2022), organisational citizenship behaviours (Li et al., 2022), work-family enrichment (Garg, 2020), psychological availability (Li et al., 2022), interpersonal helping behaviour (Kersten et al., 2021), turnover intentions (Garg et al., 2021), burnout, and stress (Saxena et al., 2020), and exhaustion and disengagement (Kersten et al., 2021). Another set of studies concluded a mediating or moderating effect of employee gratitude among the relationship between leaders' favourable feedback and employees' withdrawal behaviour (Jiang & Qu, 2022), workplace toxicity and turnover intentions (Garg et al., 2023), and corporate social responsibility and job satisfaction (Akram et al., 2023).

3.6.4 | Cluster-4: Gratitude at work, ethics, and spirituality (yellow colour)

The fourth cluster comprises seven keywords including spirituality, ethics, workplace spirituality, corporate social responsibility, and trust. Two sub-themes of studies were reported in this cluster. The first sort of studies examines the interrelationships between gratitude and workplace spirituality. Several studies have considered gratitude as a construct of workplace spirituality (Saxena et al., 2020). They argued that the Eastern conceptualisation of gratitude is incomplete without incorporation of gratitude. Besides, they elaborated that measurement of workplace spirituality is incomplete without inclusion of gratitude. Similarly, Loi and Ng (2021) also recommended that gratitude leads to enhanced experience of spirituality and meaningfulness. Büssing et al. (2021) supported and recommended gratitude as an experiential aspect of spirituality. The other sub-theme highlights different interfaces of gratitude and business ethics. Badrinarayanan et al. (2021) concluded that gratitude toward the manager moderates the relationship between ethical leadership and salespersons' emulation. In another study, Lee et al. (2022), gratitude reduces materialistic values by decreasing entitlement and perceived resource scarcity. They elaborated that gratitude-driven marketing drives reduce materialistic values and purchases and promote experiential consumption. Thus, consumers may prefer non-materialistic products and services are more effective during gratitude-related holidays (e.g., Thanksgiving, Mother's Day) Romani et al. (2013) also reported that gratitude mediates the linkage between perceived corporate social responsibility (CSR) and consumer reactions. They elaborated that firms' indulgence in CSR activities generate gratitude among beneficiary and customers leading to favourable consumer's reactions.

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4 | CONCLUSION

Employing the bibliometric analysis methods, this study offers a deeper understanding of the articles on gratitude at the workplace that has been published in Scopus-indexed journals over a period of time. Bibliometric studies help one evaluate and understand development of a field, research impact, research productivity, research trends, and future directions. Accordingly, 273 publications on gratitude at work are subjected to bibliometric analysis in the study. It revealed that publications in the field have increased steadily since 2010 and 47.25% (129 articles) have been published in the last 4 years (2018 to Feb, 2023). It suggests the growing importance and acceptance of gratitude in professional settings. Studies have suggested that gratitude is beneficial for benefactor, beneficiary, and bystander, as it develops a sense of connection and desire to reciprocate the good deeds. A good number of studies have tried to elicit gratitude among customers to reap multiple benefits like customer satisfaction, retention and loyalty, sales maximisation, and service customisation. Other studies have examined the individual and organisational impact of employees' gratitude. Individual benefits include employee well-being, commitment and engagement, psychological capital, conducive work environment, life satisfaction, trust, and compassion.

Further, the list of most influential journals is dominated by marketing journals followed by a few human resource, organisational behaviour, ethics, and spirituality journals. These results suggest that workplace gratitude studies are not confined to the domain of organisational behaviour; rather they are widely spread across diverse professional domains such as marketing, human resource, ethics, and workplace spirituality, which recommends the validity of the stakeholders' perspective of gratitude. The overlay visualisation suggests a clear preference to work upon employees' gratitude level to bring out desired results and changes in the professional settings, which is marred by workplace toxicity and a blind pursuit of materialism and success (Mehta et al., 2023). The trend is definitely toward creating organisations that offer a luxury to remain human with limitations and endless possibilities of love, trust, faith, and mutual support. Gratitude interventions like gratitude journaling, introspection, and gratitude letters help one to understand the interdependence and mutual coexistence among employees and their organisations (Bansal et al., 2023).

The country-wise analysis reports the USA as top contributor followed distantly by India, Australia, and China. It is quite understandable as the majority of major contributors of gratitude research like Roberts A Emmons, Philip Watkins, Todd B Kashdan, Sonja Lyubomirsky, Sara B Alogoe, and M E McCullough are from the USA. Also, the American researchers contributed to the early growth and development of scientific exploration of gratitude in the first decade of the twenty-first century. Now, it seems that the same trend continues in gratitude at work research with major contributions coming from the USA. The findings also report a paucity of studies by African and South American scholars. Also, studies conducted by Asian and European scholars are confined to a few countries only. It raises an immediate need of gratitude studies in different geographical and cultural settings. Such studies are poised to bring to light unexplored dimensions and effects of workplace gratitude to forefront.

The keyword co-occurrence suggests four major research areas—workplace wellbeing, customer relationship management, individual-level outcomes, and ethics and spirituality. It reflects that current focus of workplace gratitude studies are majorly confined to these four areas with emphasis on reaping its benefits. The recent studies hint toward social exchange, leader-member exchange, and HRM attribution theory. It indicates a probable amalgamation of gratitude in HR practices and policies of the companies. The social exchange states that the relationship between employer and employee is based on social exchange, wherein employees tend to respond positively to any pro-employee and empathetic move of the employer. Based on social exchange theory, it is argued that a grateful disposition by employers is going to create a positive ripple effect which will be beneficial for all stakeholders including employee, employer and the overall organisation. Similarly, leadership-exchange theory also highlights the importance of exchange/interactions between leader and follower. Garg (2020) suggested gratitude as an essential trait of an effective leader. These academic developments advocate a strong case of institutionalisation of gratitude-based interventions and practices at the workplace. Such changes have started at the college and University levels. For inference, a few universities in India have started a course on 'The art and science of gratitude' to prepare aspiring managers for future corporate journeys (Garg, 2024).

It offers several insights to future researchers to investigate numerous unexplored facets of gratitude at work. First, the current literature an acute shortage of workplace gratitude scale with majorly studies using the GQ-6, GRAT-16, or other gratitude scale (not developed in workplace contexts). Researchers have argued that workplace offers different contextual and behavioural factors to gratitude than non-professional settings. Thus, it requires scales to gauge into professionals outlook and manifestation of gratitude. Second, currently workplace gratitude studies mainly examine individual level effects with group and organisational-level effects are yet to be explored extensively. Future scholars might be interested in examining role of gratitude in facilitating group dynamics such as trust, cooperation, organisational citizenship behaviour, and pro-social activities. Similarly, gratitude could be explored in the context of organisational performance, brand image of a firm, organisational culture and climate, and organisational structures and processes. Third, existential gratitude is yet to find a place in workplace gratitude studies. Positive psychology 2.0 strongly suggests reapplication of different organisational constructs in challenging and difficult times. Thus, future studies are encourage to observe the effect of gratitude when employee or organisational are passing through turbulent time.

Despite a meticulous examination of the existing publication on gratitude at the workplace, this study has two major limitations. First, this bibliometric analysis is based on data retrieved and downloaded from the Scopus database. Although previous studies have advocated for and utilised the Scopus database, there are significant possibilities of missing articles on gratitude at work that are not indexed in the

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Scopus database. There could be publications in other databases like WoS or PsycInfo but not indexed in the Scopus database. Future researchers are advised to integrate data from different databases (after avoiding duplication of articles). It might represent a more comprehensive picture of gratitude at work research. Also, researchers could use other databases like WoS to either refute or support the findings of the present study. Second, this study is based on a relatively large sample size of 273 publications; thus, the study could not analyse the underpinning theories, dimensions, constructs, scales, interventions, and methodologies of the articles reviewed. Future researchers may conduct systematic literature review or meta-analysis studies or theme-based reviews to focus on evaluating theories, dimensions, scales or methodologies in the domain of gratitude at work. In a nutshell, the present study provides a wide-ranging encapsulation of gratitude at work research using bibliometric analysis. It offers a wide range of information on publication and citations over the years, most cited articles, most prolific authors, journals, and countries, and important research areas. Most importantly, trend analysis suggests four clusters of research works, which not only throws light on the current status of gratitude research but also foresights future directions too.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ETHICS STATEMENT

All methods in studies involving) human subjects were carried out in line with the institutional and/or national research committee's ethical standards, as well as the 1964 Helsinki Declaration and its subsequent amendments or similar ethical standards.

INFORMED CONSENT STATEMENT

Informed consent of participants was taken.

ORCID

Naval Garg https://orcid.org/0000-0001-6694-9663

REFERENCES

Aguinis, H., Dalton, D. R., Bosco, F. A., Pierce, C. A., & Dalton, C. M. (2011). Meta-analytic choices and judgment calls: Implications for theory building and testing, obtained effect sizes, and scholarly impact. *Journal of Management*, *37*(1), 5–38. https://doi.org/10.1177/0149206310377113

Akram, R., Mahmood, S., Khan, K. I., & Asghar, F. (2023). Corporate social responsibility and job satisfaction: The mediating mechanism of supervisor fairness and moderating role of gratitude. *International Journal of Business Environment*, *14*(1), 1. https://doi.org/10.1504/ijbe.2023.127689

Ali, S. B., Ahmad, M. J., Ramzan, I., Ali, M., & Khan, K. (2022). Exploring the nexus between mindfulness, gratitude, and wellbeing among youth with the mediating role of hopefulness: A South Asian perspective. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.915667

Badrinarayanan, V., Gupta, A. K., & Chaker, N. N. (2021). The pull-to-stay effect: Influence of sales managers' leadership worthiness on salesperson turnover intentions. *Journal of Personal Selling and Sales Management*, *41*(1), 39–55. https://doi.org/10.1080/08853134.2020.1820347

Bansal, S., Garg, N., & Singh, J. (2023). Exploring the psychometric properties of the Cyberbullying Attitude Scale (CBAS) and its relation with teasing and gratitude in Indian collegiates. *International Journal of Educational Management*, *37*(1), 225–239. https://doi.org/10.1108/ijem-05-2022-0198

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior. *Psychological Science*, *17*(4), 319–325. https://doi.org/10.1111/j.1467-9280.2006.01705.x

Boggiss, A. L., Consedine, N. S., Brenton-Peters, J., Hofman, P., & Serlachius, A. (2020). A systematic review of gratitude interventions: Effects on physical health and health behaviors. *Journal of Psychosomatic Research*, *135*, 110165. https://doi.org/10.1016/j.jpsychores.2020.110165

Bridger, E. K., & Wood, A. M. (2017). Gratitude mediates consumer responses to marketing communications. *European Journal of Marketing*, *51*(1), 44–64. https://doi.org/10.1108/ejm-11-2015-0810

Büssing, A., Recchia, D. R., Dienberg, T., Surzykiewicz, J., & Baumann, K. (2021). Awe/Gratitude as an experiential aspect of spirituality and its association to perceived positive changes during the COVID-19 pandemic. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.642716

Cain, I. H., Cairo, A. H., Duffy, M., Meli, L., Rye, M. S., & Worthington, E. L. (2019). Measuring gratitude at work. *The Journal of Positive Psychology*, *14*(4), 440–451. https://doi.org/10.1080/17439760.2018.1484936

Chou, C. Y., Leo, W. W. C., Tsarenko, Y., & Chen, T. (2023). When feeling good counts! Impact of consumer gratitude and life satisfaction in access-based services. *European Journal of Marketing*, *57*(2), 626–652. https://doi.org/10.1108/ejm-08-2021-0655

Davis, D. E., Choe, E., Meyers, J. D., Wade, N. G., Varjas, K., Gifford, A. Z., Quinn, A. M., Hook, J. N., Van Tongeren, D. R., Griffin, B. J., & Worthington, E. L. (2016). Thankful for the little things: A meta-analysis of gratitude interventions. *Journal of Counseling Psychology*, *63*(1), 20–31. https://doi.org/10.1037/cou0000107

DeWall, C. N., Lambert, N. M., Pond, R. S., Kashdan, T. B., & Fincham, F. D. (2012). A grateful heart is a nonviolent heart. *Social Psychological and Personality Science*, *3*(2), 232–240. https://doi.org/10.1177/1948550611416675

Di Fabio, A., Palazzeschi, L., & Bucci, O. (2017). Gratitude in organizations: A contribution for healthy organizational contexts. *Frontiers in Psychology*, *8*(2), 2025. https://doi.org/10.3389/fpsyg.2017.02025

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, *133*(1), 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070

Effendi, M., Nimran, U., Utami, H. N., & Afrianty, T. W. (2021). Effects of psychological capital and gratitude on employees intention to leave: The role of job satisfaction. *Journal of Asian Finance, Economics and Business*, *8*(2), 1125–1134. https://doi.org/10.13106/jafeb.2021.vol8.no2.1125

Fehr, R., Fulmer, A., Awtrey, E., & Miller, J. L. (2017). The grateful workplace: A multilevel model of gratitude in organizations. *Academy of Management Review*, *42*(2), 361–381. https://doi.org/10.5465/amr.2014.0374

Garg, N. (2017). Workplace spirituality and organizational performance in Indian context: mediating effect of organizational commitment, work motivation and employee engagement. *South Asian Journal of Human Resources Management*, *4*(2), 191–211. https://doi.org/10.1177/2322093717736134

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Garg, N. (2020). Exploring role of gratitude in developing teacher leadership in Indian universities. *International Journal of Educational Management*, *34*(5), 881–901. https://doi.org/10.1108/ijem-07-2019-0253

Garg, N. (2024). Gratitude research: Review and future agenda using bibliometric analysis of the studies published in the last 20 years. *Asian Journal of Social Psychology*. https://doi.org/10.1111/ajsp.12621

Garg, N., Gera, S., & Punia, B. K. (2021). Gratitude and turnover intention in E-Commerce startups. *Journal of Electronic Commerce in Organizations*, *19*(2), 34–54. https://doi.org/10.4018/jeco.2021040103

Garg, N., Jain, A., & Punia, B. (2020). Gratitude, social intelligence, and leadership among university teachers: Mediation and moderation analysis. *International Journal of Organizational Analysis*, *29*(2), 368–388. https://doi.org/10.1108/ijoa-01-2020-2016

Garg, N., Kumari, S., & Punia, B. (2022a). Resolving stress of university teacher: Exploring role of workplace spirituality and constructive workplace deviance. *South Asian Journal of Business Studies*, *11*(3), 295–315. https://doi.org/10.1108/sajbs-04-2020-0118

Garg, N., Kumari, S., & Punia, B. K. (2022b). Interrelation between gratitude and vitality among students: Exploring mediating role of resilience. *Psychological Studies/Psychological Studies*, *67*(2), 200–207. https://doi.org/10.1007/s12646-022-00654-x

Garg, N., Mahipalan, M., Poulose, S., & Burgess, J. (2022c). Does gratitude ensure workplace happiness among university teachers? Examining the role of social and psychological capital and spiritual climate. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.849412

Garg, N., Verma, S., & Palframan, J. T. (2023). Positive reframing as a mediator between gratitude and technostress amongst Indian students during the COVID-19 pandemic. *Kybernetes*, *52*(8), 2810–2825. https://doi.org/10.1108/k-12-2021-1250

Guzzo, R. F., Wang, X., & Abbott, J. (2020). Corporate social responsibility and individual outcomes: The mediating role of gratitude and compassion at work. *Cornell Hospitality Quarterly*, *63*(3), 350–368. https://doi.org/10.1177/1938965520981069

Hameed, A., & Khwaja, M. G. (2022). The impact of benevolent human resource management attributions on employees' general work stress, with the mediating influence of gratitude. *Journal of General Management*, 30630702211308. https://doi.org/10.1177/03063070221130872

Huang, M. (2015). The influence of relationship marketing investments on customer gratitude in retailing. *Journal of Business Research*, *68*(6), 1318–1323. https://doi.org/10.1016/j.jbusres.2014.12.001

Jiang, Z., & Qu, Y. (2022). Leader favorable feedback and withdrawal behavior: A moderated mediation model of gratitude and leader-member exchange. *Personnel Review*, *52*, 972–991. https://doi.org/10.1108/pr-04-2021-0228

Kersten, A., Van Woerkom, M., Kooij, D., & Bauwens, R. (2021). Paying gratitude forward at work. *Journal of Personnel Psychology*, *21*(3), 137–148. https://doi.org/10.1027/1866-5888/a000296

Lee, H., Chugani, S., & Namkoong, J. (2022). The role of entitlement and perceived resources in gratitude's effect on materialism: Longitudinal and situational effects. *Journal of Business Research*, *139*, 993–1003. https://doi.org/10.1016/j.jbusres.2021.10.029

Li, N., Zheng, X., Yu, Y., & Yu, J. (2022). A quasi-experimental examination of knowledge-sharing interventions enhancing service performance: The roles of time, occupational identity, and image. *Journal of Organizational Behavior*, *43*, 818–839. https://doi.org/10.1002/job.2609

Liang, J., & Guo, L. (2021). Gratitude and sustainable consumer behavior: A moderated mediation model of time discounting and connectedness to the future self. *Psychology & Marketing*, *38*(8), 1238–1249. https://doi.org/10.1002/mar.21502

Liao, Z., Wu, L., Zhang, H., Song, Z., & Wang, Y. (2022). Exchange through emoting: An emotional model of leader–member resource exchanges. *Personnel Psychology*, *76*(1), 311–346. https://doi.org/10.1111/peps.12506

Liu, C., Kuo, T. M., Wang, Y., Shen, Y., Chen, S., & Hong, J. (2022). Perceived luxurious values and pay a price premium for Michelin-starred restaurants: A sequential mediation model with self-expansion and customer gratitude. *International Journal of Hospitality Management*, *103*, 103185. https://doi.org/10.1016/j.ijhm.2022.103185

Locklear, L. R., Sheridan, S., & Kong, D. T. (2022). Appreciating social science research on gratitude: An integrative review for organizational scholarship on gratitude in the workplace. *Journal of Organizational Behavior*, *44*(2), 225–260. https://doi.org/10.1002/job.2624

Loi, N. M., & Ng, D. H. (2021). The relationship between gratitude, well-being, spirituality, and experiencing meaningful work. *Psych*, *3*(2), 85–95. https://doi.org/10.3390/psych3020009

Mangus, S. M., Bock, D. E., Folse, J. A. G., & Jones, E. (2022). The comparative effects of gratitude and indebtedness in B2B relationships. *Industrial Marketing Management*, *103*, 73–82. https://doi.org/10.1016/j.indmarman.2022.03.009

McCullough, M. E., Emmons, R. A., & Tsang, J. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*, *82*(1), 112–127. https://doi.org/10.1037/0022-3514.82.1.112

Mehta, P., Garg, N., Gharib, M., Mehak, N., & Pimpunchat, B. (2023). Social support during COVID-19: Exploring the psychometric properties of the PSS-JSAS and its relationship with job search activities. *Socio-economic Planning Sciences*, *85*, 101349. https://doi.org/10.1016/j.seps.2022.101349

Mittal, S., & Maity, M. (2022). Is relationship marketing investments–customer gratitude–customer loyalty linkage influenced by contextual and individual moderators? An emerging economy perspective. *Journal of Indian Business Research*, *14*(4), 492–519. https://doi.org/10.1108/jibr-11-2021-0375

Morales, A. C. (2005). Giving firms an "e" for effort: Consumer responses to high-effort firms. *Journal of Consumer Research*, *31*(4), 806–812. https://doi.org/10.1086/426615

Ni, D., Song, L. J., Zheng, X., Zhu, J., Zhang, M., & Xu, L. (2022). Extending a helping hand: How receiving gratitude makes a difference in employee performance during a crisis. *Journal of Business Research*, *149*, 967–982. https://doi.org/10.1016/j.jbusres.2022.05.055

Park, J. C., Choi, J., & Yeu, M. (2016). Relationship between corporate philanthropy and consumer loyalty-the mediating role of gratitude, trust and commitment: South Korean consumer perspectives. *Academy of Marketing Studies Journal*, *20*(1), 1. https://www.questia.com/library/journal/1G1-459230035/relationship-between-corporate-philanthropy-and-consumer

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation*, *25*, 348. http://ci.nii.ac.jp/naid/10016754266

Qi, J. M., Peng, Y., Lowman, G. H., & He, X. (2023). The impact of service climate on gratitude in driving customer outcomes. *Journal of Services Marketing*, *37*(1), 78–95. https://doi.org/10.1108/jsm-12-2021-0458

Rasool, S. F., Wang, M., Tang, M., Saeed, A., & Iqbal, J. (2021). How toxic workplace environment effects the employee engagement: The mediating role of organizational support and employee wellbeing. *International Journal of Environmental Research and Public Health*, *18*(5), 2294. https://doi.org/10.3390/ijerph18052294

Ren, L., Liu, Y., & Yin, Y. (2023). Do grateful employees take charge more in China? A joint moderating effect model. *Asia Pacific Business Review*, *29*(1), 70–88. https://doi.org/10.1080/13602381.2022.2120252

Romani, S., Grappi, S., & Bagozzi, R. P. (2013). Explaining consumer reactions to corporate social responsibility: The role of gratitude and altruistic values. *Journal of Business Ethics*, *114*(2), 193–206. https://doi.org/10.1007/s10551-012-1337-z

Sarkar, A., Garg, N., Srivastava, D., & Punia, B. (2023). Can gratitude counter workplace toxicity? Exploring the mediating role of psychological capital (Psy CaP). *Business Perspectives and Research*, *12*(2), 261–276. https://doi.org/10.1177/22785337221148300

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Saxena, A., Garg, N., Punia, B., & Prasad, A. (2020). Exploring role of Indian workplace spirituality in stress management: A study of oil and gas industry. *Journal of Organizational Change Management/Journal of Organisational Change Management*, *33*(5), 779–803. https://doi.org/10.1108/jocm-11-2019-0327

Schnall, S., & Cannon, P. R. (2012). The clean conscience at work: Emotions, intuitions and morality. *Journal of Management, Spirituality & Religion*, *9*(4), 295–315. https://doi.org/10.1080/14766086.2012.742749

Septianto, F., Kemper, J. A., & Northey, G. (2020). Thanks, but no thanks: The influence of gratitude on consumer awareness of food waste. *Journal of Cleaner Production*, *258*(1), 120591. https://doi.org/10.1016/j.jclepro.2020.120591

Simon, F., & Tossan, V. (2018). Does brand-consumer social sharing matter? A relational framework of customer engagement to brand-hosted social media. *Journal of Business Research*, *85*, 175–184. https://doi.org/10.1016/j.jbusres.2017.12.050

Steinhoff, L., & Palmatier, R. W. (2014). Understanding loyalty program effectiveness: Managing target and bystander effects. *Journal of the Academy of Marketing Science*, *44*(1), 88–107. https://doi.org/10.1007/s11747-014-0405-6

Suetrong, P., & Pires, G. D. (2021). Brand liking and consumer gratitude effects on consumer-brand love relationships and repurchase intention. *International Journal of Economics and Business Research*, *21*(4), 458. https://doi.org/10.1504/ijebr.2021.115504

Van Dierendonck, D., & Patterson, K. (2014). Compassionate love as a cornerstone of servant leadership: An integration of previous theorizing and research. *Journal of Business Ethics*, *128*(1), 119–131. https://doi.org/10.1007/s10551-014-2085-z

Wetzel, H. A., Hammerschmidt, M., & Zablah, A. R. (2014). Gratitude versus entitlement: A dual process model of the profitability implications of customer prioritization. *Journal of Marketing*, *78*(2), 1–19. https://doi.org/10.1509/jm.12.0167

Zhan, X., Yang, W., Guo, Y., & Luo, W. (2022). Can patient gratitude compensate for depletion from family-to-work conflict in nurses? An experience sampling study. *Personnel Review*, *52*(9), 2153–2177. https://doi.org/10.1108/pr-12-2020-0891

Zhao, T., Li, H., Zheng, L., & Zhang, Y. (2023). How dispositional gratitude shapes employee well-being and organizational commitment: The mediating roles of leader-member exchange and coworker exchange. *Journal of Career Assessment*, *31*(1), 149–171. https://doi.org/10.1177/10690727221099867

AUTHOR BIOGRAPHIES

Dr. Naval Garg is a dynamic researcher and academician who has made significant contributions to the field of Organisational Behaviour and Human Resource Management. Currently serving as an Assistant Professor at the University School of Management and Entrepreneurship, DTU, his industry experience is of more than 5 years. Dr. Garg has authored more than 70 research papers and has served as the editors of various books and has an h-index of 20. He is also listed as reviewer at various esteemed journals of SSCI, SCI, ABDC and Scopus indexing. His research interests include various aspects of organization behaviour, workplace spirituality, nonviolent work behaviour and psychometric properties analysis.

Prof. B.K. Punia Having worked as Vice Chancellor, Dean, Director and Head at various institutions, Prof. B.K. Punia is currently working as Professor (w.e.f. March, 2005) at Haryana School of Business, Guru Jambheshwar University of Science & Technology, Hisar. Also, he is a renowed researcher with research papers in various international journals of repute.

How to cite this article: Garg, N., & Punia, B. K. (2024). Workplace gratitude studies: Review and future agenda using bibliometric analysis. *Journal of Public Affairs*, *24*(3), e2934. https://doi.org/10.1002/pa.2934

14791854, 2024, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pa.2934 by Delhi Technological University, Wiley Online Library on [10/07/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License